Kant's Metaphysics of Mind and Rational Psychology

DISSERTATION
zur Erlangung des akademischen Grades

Doctor philosophiae
(Dr. phil.)

eingereicht

an der Philosophischen Fakultät I
der Humboldt-Universität zu Berlin

von Dr. Steven Tester

Die Präsidentin/Der Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Die Dekanin/Der Dekan der Philosophischen Fakultät I
Prof. Dr. Michael Seadle

Gutachter/innen

Erstgutachter/in: Prof. Dr. Rolf-Peter Horstmann

Zweitgutachter/in: Dr. Rachel Zuckert

Tag der Disputation: August 20, 2014

Kant's Metaphysics of Mind and Rational Psychology

Abstract

This dissertation considers Kant's discussions of the metaphysics of mind in his critical encounter with the rational psychology of Baumgarten, Wolff, and others in the *Critique of Pure Reason* and his lectures on metaphysics. In contrast with prevailing interpretations, I argue that Kant does not offer a straightforward rejection of his predecessors but that he retains some commitments to the substantial view of the self and modifies others within the framework of transcendental idealism to provide accounts of the nature of personhood, mental powers, the possibility of mind-body interaction, and the possibility of freedom of the will. This interpretation of Kant reveals continuity between Kant's pre-critical and critical positions on the metaphysics of mind and points forward to a role for aspects of Kant's metaphysics of mind in his practical philosophy.

Die Dissertation diskutiert die kantische Metaphysik des Geistes anhand der in der *Kritik der reinen Vernunft* und den aus dem Nachlass veröffentlichten Vorlesungen zur Metaphysik geleisteten Auseinandersetzung mit der rationalen Psychologie seiner Vorgänger, insbesondere Baumgarten und Wolff. Es wird dafür argumentiert, dass Kant die Meinungen seiner Vorgänger nicht uneingeschränkt zurückweist, sondern die Vorstellung der Seele als Substanz in seine Diskussion der Personalität, mentaler Kräfte, der Möglichkeit einer Körper-Seele Interaktion sowie der Willensfreiheit teilweise beibehält. Ein Verdienst dieser Interpretation ist es, die Kontinuität zwischen Kants vorkritischer Position und seiner kritischen Philosophie aufzuzeigen. Darüber hinaus soll aber auch auf eine wichtige Funktion der kantischen Metaphysik des Geistes für seine praktische Philosophie hingewiesen werden.

Acknowledgements

I owe a debt of gratitude to a number of people and institutions for their support during this project. I am grateful to Rolf-Peter Horstmann for his willingness to advise me and for his comments and criticism at various stages. And I am especially appreciative of Rachel Zuckert for her extensive and insightful comments on drafts of chapters from this dissertation on Kant as well as my other dissertation on Lichtenberg at Northwestern University. I would also like to thank the particpants in Rolf-Peter Horstmann's colloquium and Tobias Rosefeldt's colloquium at Humboldt Universität where I had the opportunity to present my work in progress over the past few years. I have also benefited at various points from discussions with Lucy Allais, Catherine Diehl, Corey W. Dyck, Wolfgang Ertl, Peter Fenves, Andree Hahmann, Till Hoeppner, Bernd Ludwig, Colin Marshall, Colin Mclear, James Messina, Tyke Nunez, Tobias Rosefeldt, B. Scot Rousse, Karl Schafer, Karsten Schoellner, Henry Southgate, Nicholas Stang, Bernard Thöle, and Falk Wunderlich among others. The project was also made possible with generous funding from the Elsa-Neumann-Stipendium des Landes Berlin and the administrative assistance of Sabine Hassel. Finally I would like to thank my family for their encouragement and emotional support: my parents John and Anna Tester, my brother Danny, my wife Karen Carolin, and my children Vivian and Elliot.

Contents

Introduction

What are the faculties of the mind? How do they contribute to the unity of thought? Must they be grounded in a simple substance? What is the relationship between the substance that grounds thoughts and personhood? How is mind-body interaction possible? Do we have freedom of the will? These are a series of questions with which German philosophers of the seventeenth and eighteenth century struggled in their pursuit of a coherent rational psychology. For the German rationalists such as Wolff and Baumgarten who followed Leibniz the answer to these questions regarding the metaphysics of mind and freedom lay in an understanding of the ontological ground of thought as a substance.[1] According to a shared basic view among the German rationalists, the soul is an immaterial substance that serves as the ontological ground for mental capacities and thought. This substance is conceived of as a power or as possessing a power that is the sufficient ground of the inherence of the attributes of thought in the substance and a sufficient ground for effecting changes in itself and in surrounding objects. The immaterial, substantial ground of thought is also held to be simple, and therefore immortal, and capable of free actions. Although the German rationalists shared this basic view of rational psychology and the metaphysics of mind, there were nevertheless serious debates among them regarding the details of how this position could be argued for and what it might entail. For example, Christian Wolff and Alexander Baumgarten disagreed about whether the substantiality of the soul was something that could be established through a posteriori empirical observation of thought or whether the only promising route to establish substantiality was through an a priori argument. There were also disagreements between Wolff and his supporters and Pietist philosophers such as Christian August Crusius. Wolff and Crusius disagreed, for example, about whether the mental faculties that make the unity of thought possible could be reduced to a single power of a substance or whether one might accept a plurality of powers. Kant's predecessors also disagreed about fundamental issues such as the nature of mental causation and mind-body interaction, with some philosophers arguing for physical influx while others adhered more closely to the Leibnizian doctrine of

[1] For a discussion of the philosophy of Kant's predecessors in Germany, particularly Wolff and Crusius, see Lewis White Beck, *Early German Philosophy* (Cambridge: Harvard University Press, 1969), pp. 256–305 and 393–401.

pre-established harmony. Many of these discussions also took place in response to the specter of the materialist conception of mind.[2]

Although the details of many of these debates have been obscured by history, Kant in his time was intimately familiar with the developments in the metaphysics of mind in rational psychology and was an engaged participant in a number of the debates. In his writings from 1747 and 1781 preceding the publication of the *Critique of Pure Reason*, Kant pursues questions related to the metaphysics of mind assuming an ontology involving substantial monads similar to the Leibnizian ontology shared by the rationalists and a methodology that is primarily rationalist or a priori. For example, in *Thoughts on the True Estimation of Living Forces* (1747) Kant argues for a single power or force that makes mental causation and mind-body interaction possible, and in the *New Elucidation of the First Principles of Metaphysical Cognition* (1755) Kant argues against pre-established harmony and in favor of a realist view of substantial causal interaction that appeals to the existence of God as the guarantor of genuine interaction among mental and physical substances. However, in the lectures on metaphysics in the 1770's leading to the publication of the *Critique of Pure Reason* in 1781, and in Kant's critical period, there is a marked shift in the standpoint and the methodology from which Kant considers questions regarding the metaphysics of mind. The shift is particularly evident in *On the Form and Principles of the Sensible and Intelligible World* (1770), where Kant begins to develop his doctrine of transcendental idealism arguing that space and time are merely subjective a priori intuitions and "not something objective and real" (AA 2:400), which forms the core of his mature philosophical system.[3]

---

[2] On materialism in the eighteenth century, see John W. Yolton, *Thinking Matter: Materialism in Eighteenth–Century Britain* (Oxford: Blackwell, 1984), and Yolton, *Locke and French Materialism* (Oxford: Clarendon Press, 1991).

[3] All references to Kant are to the *Akademie Ausgabe*: Immanuel Kant, *Kant's gesammelte Schriften*, ed. Preussische Akademie der Wissenschaften and Deutsche Akademie der Wissenschaften zu Berlin (Berlin: De Gruyter, 1900–). The *Critique of Pure Reason* is cited according to the standard A/B edition and page number, and other works are cited according to volume and page (e.g. AA x:xx). Unless otherwise noted, translations are from: Immanuel Kant, *Critique of Pure Reason*, ed. and trans. Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1998); *Anthropology From A Pragmatic Point of View*, ed. and trans. Robert B. Louden (Cambridge: Cambridge University Press, 2006); *Lectures on Metaphysics*, ed. and trans. Karl Ameriks and Steve Naragon (Cambridge: Cambridge University Press, 1997); *Theoretical Philosophy 1755-1770*, ed. and trans. D. Walford with R. Meerbote (Cambridge: Cambridge University Press, 1992); *Theoretical Philosophy after 1781*, ed. H. Allison, P. Heath and trans. G. Hatfield, M. Friedman, H. Allison, P. Heath (Cambridge: Cambridge University Press, 2002).

As Kant's thinking on transcendental idealism matures into the doctrine found in the *Critique of Pure Reason*, he begins to raise questions regarding the metaphysics of mind within the conceptual framework of transcendental idealism and the distinction between appearances and things in themselves. Rather than directly considering whether, for example, the unity of thought can be grounded in a composite of substances each endowed with a cognitive power or considering whether the cognitive powers can be reduced to a single power, Kant formulates his questions in terms of the critical philosophy and its commitment to the ideality of space and time. Kant wonders, for example, what the ground of the unity of consciousness must be like given the fact that we appear to ourselves as spatial and temporal objects and appearances are grounded in the non-spatial and non-temporal way things are in themselves. Or rather than attempting to explain mind-body interaction on the basis of a direct appeal to fundamental powers or forces as he had done in the *Living Forces* essay, Kant considers why philosophers have held that mind and body are heterogeneous things in themselves rather than appearances and how this misunderstanding has led to problematic solutions to the problem of mind-body interaction. This shift is nowhere more evident than in Kant's discussions of the metaphysics of mind in the Paralogisms of Pure Reason and the Third Antinomy, which are both focused on how transcendental idealism can contribute to an understanding of and resolution to some of the major problems that faced rational psychology in discussing the soul and its various properties.

However, interpreters have often understood Kant to be providing only a negative criticism of rational psychology in the *Critique of Pure Reason* on the basis of the critical epistemology of transcendental idealism and its claim that we can cognize things only as they appear to us given our spatial and temporal forms of intuition and not as they are in themselves rather than a positive contribution to metaphysical debates about the metaphysics of mind in light of the distinction between appearances and things in themselves. There are perhaps several reasons why Kant has tended to be interpreted this way. Often, Kant's claim that in criticizing rational psychology he is not criticizing any particular doctrine but a mere tendency of reason is seen as a reason to look no further into the context within which Kant was developing his views. And where there has been a willingness among interpreters to see Kant as responding to particular predecessors, the target is generally uncritically taken to be either Descartes in the case of the Paralogisms or Hume in the case of the Transcendental

Deduction.[4] The tendency to overlook the rationalist background of Kant's discussion and the profound influence of rationalism upon his thought have led interpreters to misunderstand the actual arguments that Kant is targeting in his discussion of rational psychology and more importantly to overlook aspects of Kant's discussion that reveal a commitment to positive metaphysical doctrines very similar to those proposed by the rational psychologists and continuous with the rationalism of his pre-critical writings. In addition to overlooking Kant's close connections with rationalism, there has also often been a tendency to take Kant's claim that his pre-critical writings are anathema to the critical project and should be ignored too seriously, and so interpreters often see little continuity between the metaphysics of the mind in the pre-critical period and Kant's discussions of rationalism and the metaphysics of the mind in the *Critique of Pure Reason*.

It has also been easier to overlook Kant's positive proposals regarding the metaphysics of the mind because of widespread adherence to an epistemological interpretation of Kant's transcendental idealism and his distinction between appearances and things in themselves.[5] By emphasizing Kant's claims about the epistemic inaccessibility of the soul as the things in itself that grounds thought, interpreters have held that Kant's aim was to show how the rationalist is misled into positing certain properties of the soul because they mistake the transcendental unity of apperception, which is accessible to us, with the ground of the unity of thought, which is not accessible to us.[6] A similar approach is taken to explain Kant's criticism of rationalist ideas of personhood and mental causation.[7] Kant's statements that appear more metaphysical and rationalist in tenor have simply been interpreted as an anomaly or as confusion or overreaching on Kant's part. I will argue, however, that if Kant's

---

[4] See, for example: Jonathan Bennett's discussion of Kant's criticism of Descartes in *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974), pp. 66–81, Patricia Kitcher's discussion of Kant's criticism of Hume's view of the self and personal identity in "Kant on Self-Identity," *Philosophical Review* 91(1) (1982), pp. 41–72, and Kitcher, "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), pp. 515–547.

[5] For representative epistemological interpretations of Kant's transcendental idealism, see: Henry Allison, *Kant's Transcendental Idealism. An Interpretation and Defense* (New Haven: Yale University Press, 1983); Gerold Prauss, *Kant und das Problem der Dinge an Sich* (Bonn: Bouvier, 1974).

[6] For epistemologically-oriented interpretations of the Paralogisms, see: Patricia Kitcher, "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), pp. 515–547; Michelle Grier, "Illusion and Fallacy in Kant's First Paralogism," *Kant-Studien*, 84(3) (1993), pp. 257–282.

[7] See, for example: C. Thomas Powell, "Kant's Fourth Paralogism," *Philosophy and Phenomenological Research* 48(3) (1988), pp. 389–414, and Powell, *Kant's Theory of Self-Consciousness* (Oxford: Oxford University Press, 1990).

distinction between appearances and things in themselves is regarded as an ontological one, then it becomes clearer that his discussion of the metaphysics of mind does not stop with a confession of ignorance and a criticism of rationalist hubris. Rather, one sees, particularly in the first edition of the *Critique of Pure Reason* (1781), that Kant engages directly with a number of debates concerning the metaphysics of mind in rationalist psychology and provides arguments that enlist the distinction between appearances and things in themselves to make positive claims about the substantiality, simplicity, and persistence of the ground of thought as well as positive contributions to debates about personhood, mind-body interaction, and the possibility of freedom of the will.

This is not to say, however, that in pointing out Kant's indebtedness to his rationalist predecessors and interpreting his views as metaphysical rather than merely critical that I am alone in doing so. Indeed, there has been some excellent work in this area in recent years. Recently Corey W. Dyck and Falk Wunderlich have done a great deal to expand our knowledge of Kant's relationship to his rationalist predecessors on the topics of mind and cognition.[8] Dyck's work, for example, has focused in part on Kant's criticism of Wolff's attempt to ground the claims of rational psychology on empirical psychology. However, as important and interesting as this research has been, it does neglect aspects of Kant's discussion of the rationalists in which he is not raising epistemological objections to rationalist claims about knowledge of the soul but engages in a priori metaphysics about the soul or substance as a ground of thought. Eric Watkins has also done a great deal to show that Kant's account of causality in the *Critique of Pure Reason* is in some regards continuous with his pre-critical views and borrows a great deal from the rationalists.[9] However, Watkins's account focuses more generally on Kant's account of substantial causality and less so on how Kant's physical-influx account of mind-body interaction fits within the broader framework of Kant's discussion of rational psychology and its concerns with substantiality, mental powers, personhood, and freedom in the *Critique of Pure Reason*. There has also been a great deal of work recently by Lucy Allais, Rae Langton, Tobias Rosefeldt and others who argue that Kant's distinction between appearances and things in themselves should be understood as an

---

[8] See: Corey W. Dyck, "The Divorce of Reason and Experience: Kant's Paralogisms of Pure Reason in Context," *Journal of the History of Philosophy* 47(2) (2009), pp. 249–275; Falk Wunderlich, "Kant's Second Paralogism in Context," in *Between Leibniz, Newton and Kant*, ed. W. Lefevre (Netherlands: Springer, 2001), pp. 175–188.

[9] See Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005).

ontological one, which provides a great deal of support for my approach to understanding the role transcendental idealism plays in Kant's confrontation with the rationalists.[10] In arguing that things in themselves are substances in some regard, my approach also aligns with a tradition of Kant interpretation that includes earlier treatments by Heinz Heimsoeth and Max Wundt of Kant as a metaphysician.[11]

As I will argue, the positive views on the metaphysics of mind that Kant develops in confrontation with his rationalist predecessors, and surrounding issues concerning the unity of thought, the nature of personhood, mental causation, and freedom of the will are also more than a mere anomaly in Kant's theoretical philosophy but are integral to his later considerations about the nature of moral responsibility, agency, and duties in his practical philosophy. The rationalists to whom Kant is responding aimed to establish the conclusion that the soul is a simple, immaterial substance in order to demonstrate its immortality in the afterlife. As a simple substance, the soul would not be subject to corruption and so would be susceptible to just rewards and punishment in the afterlife. As Kant says, one "final aim to which in the end the speculation of reason in its transcendental use is directed" is "the immortality of the soul" (A 798/B 826). Kant argues, however, that the rationalist is unable to establish the substantiality of the soul in a way that would secure its persistence and its immortality. In contrast with the aims of the rationalist to secure the soul's immortality, Kant argues that the desire of reason to go beyond the boundaries of knowledge and to speculate about metaphysical questions regarding the unknowable ground of thought is motivated by a practical interest. And in doing so, he makes the ultimate aim of the consideration of the metaphysics of mind for ethics more explicit than his rationalist predecessors had done. For Kant, the practical interest of reason is in securing an understanding of the metaphysics of mind that is necessary for the purposes of ethics and the assessment of moral responsibility. Kant explicitly mentions the need to secure the possibility of freedom in the Third Antinomy, but it is clear that more issues concerning the metaphysics of mind are at stake in his

---

[10] See: Lucy Allais, "Kant's One World," *British Journal for the History of Philosophy* 12(4) (2004), pp. 655–684; Tobias Rosefeldt, "Dinge an sich und sekundäre Qualitäten," in *Kant in der Gegenwart*, ed. J. Stolzenburg (Berlin: de Gruyter, 2007), pp. 167–209; Rae Langton, *Kantian Humility: Our Ignorance of Things in Themselves* (Oxford: Oxford University Press, 1998).

[11] See: Heinz Heimsoeth, *Studien zur Philosophie Immanuel Kants*: *Metaphysische Ursprünge und Ontologische Grundlagen* (Köln: Kölner Universitäts Verlag, 1956); Max Wundt, *Kant als Metaphysiker–Ein Beitrag zur Geschichte der deutschen Philosophie im 18. Jahrhundert* (Stuttgart: Ferdinand Enke, 1924).

discussion of rational psychology.[12] The possibility of freedom is a necessary condition for the imputability of actions to persons and so also for the assessment of moral responsibility. However, imputability also requires an understanding of whether the person who undertook some action is the same as the one to whom the actions are imputed. And each of these questions is intimately linked with questions about whether the ground of thought is a substance, whether mind-body interaction is possible, and what kinds of mental powers a person has. For Kant, such considerations regarding the necessary conditions for imputability and moral responsibility in ethics require much more than an epistemologically motivated critique of rational psychology and a rejection of its central doctrines on the substantiality of the soul. They require a sustained consideration of the metaphysics of mind in light of the doctrine of transcendental idealism. As such, one might see Kant's considerations of the metaphysics of mind as providing a kind of groundwork for his later views on morality and practical philosophy.

In what follows, I argue that a close consideration of the arguments of the rational psychologists and Kant's discussion of these arguments in the light of his distinction between appearances and things in themselves, particularly in the first edition of the *Critique of Pure Reason*, reveals that Kant proposed positive metaphysical views regarding the grounds of thought and their relationship to mental powers, personhood, the possibility of mind-body interaction, and the possibility of freedom of the will. Kant maintains that we have certain mental capacities – understanding, sensibility, and reason – that are necessary for the unity of thought and for objective judgments about the world, and we have these capacities in virtue of our powers of spontaneity and receptivity that are grounded in substances. This picture of the metaphysics of mind also underlies Kant's positive views on the nature of personhood, the possibility of mind-body interaction, and the possibility of freedom of the will. Personhood consists in the persistence of our mental capacities whereby we unify our thoughts. Interaction is possible through the powers of the substances that ground mental and physical appearances. And freedom of the will is possible through the power of spontaneity that allows us to reason.

My approach to interpreting Kant's views on the metaphysics of mind and their relationship to rational psychology aims not only to be attentive to the "letter" of Kant's writings but also to the "spirit" of his philosophy. This is to say that I am attentive to the wide range of texts in which Kant expresses his views, including the *Critique of Pure Reason* and

---

[12] See A 802f./ B830f.

his lectures and *Reflexionen*, although I also rationally reconstruct portions of Kant's arguments and engage with contemporary interpretive debates. By contextualizing Kant's views, I also hope to provide a framework within which to give his views a plausible, historically sensitive interpretation. With the exception of occasional appeals to the second edition *Critique of Pure Reason* (1787), my discussion focuses primarily on the first edition Paralogisms of Pure Reason, Transcendental Deduction, and the Third Antinomy because of the strong resources these sections provide for understanding Kant's commitments to the existence of things in themselves as grounds of appearances and the influence of rational psychology on Kant's metaphysics of mind. Although there may be a shift in Kant's thinking regarding the ontological ground of thought in the second edition of the *Critique of Pure Reason*, I will not consider here whether such a shift indeed occurs or what Kant's motivations might have been for the shift.[13] Nor has my aim been to provide a decisive resolution to any longstanding debates about Kant's seemingly inconsistent claims regarding positive features of things in themselves and his claim that we have no cognition of things as they are in themselves. Such questions will need to be addressed at a later time within a more detailed consideration of the scope and aims of Kant's transcendental idealism as a whole.

I argue for this interpretation of Kant's metaphysics of mind in the following way. The first chapter (1) considers Kant's views on whether the ground of the attributes of thought is a substance. In contrast with Corey W. Dyck who argues that Kant's primary target in the first Paralogism is a Wolffian argument for the substantiality of the soul, I argue that Kant's primary target in the First Paralogism is Baumgarten's a priori argument for the substantiality of the soul, which suggests that the substantiality of the soul as the absolute ground of thought follows from the principle of sufficient reason and the fact that thought is an attribute.[14] If we understand Kant as responding to Baumgarten's a priori argument for substantiality, then we are also in a position to resolve the longstanding interpretive problem of understanding Kant's claim that the rationalist's argument is motivated by "transcendental illusion."[15] The rationalist's illusion consists in thinking that the application of the principle

---

[13] On the difference between Kant's views on the self in the A and B edition Paralogisms, see Rolf-Peter Horstmann, "Kants Paralogismen," *Kant-Studien* 84(4) (1993), pp. 408–425.

[14] Corey W. Dyck, "The Divorce of Reason and Experience: Kant's Paralogisms of Pure Reason in Context," *Journal of the History of Philosophy* 47(2) (2009), pp. 249–275.

[15] On transcendental illusion and the Paralogisms, see Patricia Kitcher, "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), p. 518; Jonathan Bennett, *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974), p. 281. Michelle Gilmore Grier, "Illusion and Fallacy in

of sufficient reason will lead to the conclusion that the ultimate substantial ground of thought is an appearance. Since this substance is understood as an appearance, it is thought to be a persisting and abiding substance. Kant argues in contrast that the ultimate ground posited by the principle of sufficient reason cannot be an appearance but must be a thing in itself. Since, however, it is a thing in itself, the spatial and temporal conception of substance does not apply to it. And since this conception of substance does not apply, the rationalist cannot conclude that the ultimate ground of thought is a persisting, immortal substance. However, although Kant rejects the idea that the ultimate ground of thought is a persisting substance, I also show that he nevertheless appears to maintain that there is a substance that grounds thought. In contrast with a recent interpretation by Julian Wuerth, which also holds that Kant was committed to a substance as the ground of thought, I argue that such a substance cannot be merely a bare substratum but must have intrinsic powers that allow it to ground thought.[16] And I defend this interpretation of substance against Langton's claim that for Kant powers cannot be intrinsic properties of a substance.[17]

Having shown that Kant may have held that a substantial thing in itself grounds thought through its powers, I turn in the second chapter (2) to a consideration of Kant's discussion of debates among rationalists regarding the number of powers the soul may possess. Wolff and Wolffian philosophers argue that the soul may possess only a single power of representation, a *vis repraesentativa*, otherwise it would be a composite and therefore subject to perishing through the dissolution of its parts. Although it is widely recognized by Henrich, Dyck and others that Kant considers the number of our mental powers in the subjective deduction, I argue that Kant also implicitly treats this question in the discussion of the Second Paralogism.[18] In the subjective deduction, Kant argues that we have multiple irreducible mental powers. However, Kant's claim that the mind has multiple powers leaves him open to the Wolffian objection that the existence of multiple mental powers entails that the soul is a composite. I show that in the Second Paralogism, Kant

---

Kant's First Paralogism," *Kant-Studien* 84(3) (1993), p. 258; Ian Proops, "Kant's First Paralogism," *Philosophical Review* 119(4) (2010), pp. 449–495.

[16] See Julian Wuerth, "The First Paralogism, its Origin, and its Evolution: Kant on How the Soul Both Is and Is Not a Substance, " in *Cultivating Personhood: Kant and Asian Philosophy*, ed. Stephen R. Palmquist (Berlin: de Gruyter, 2010), pp. 157–166.

[17] See Rae Langton, *Kantian Humility: Our Ignorance of Things in Themselves* (Oxford: Oxford University Press, 1998), p. 118.

[18] See: Corey W. Dyck, "The Subjective Deduction and the Search for a Fundamental Force," *Kant-Studien* 99(2) (2008), pp. 152–179; Dieter Henrich, "On the Unity of Subjectivity," in *The Unity of Reason* (Cambridge: Harvard University Press, 1994), pp. 19–40.

provides an argument that shows that compositeness, in the sense that concerns the Wolffians, applies only to appearances and not to things in themselves. The argument also suggests an argument against the Wolffian claim that the existence of multiple powers entails a composite soul. Because the soul as the ultimate ground of the attributes of thought is not an appearance, it may possess multiple powers and nevertheless not be a composite. This Kantian argument extends arguments made by Crusius and Lange against the Wolffian doctrine of a single power and its construal of powers mereologically as parts of the soul occupying distinct regions of space. Although Kant shows that the existence of multiple powers is compatible with their inherence in a simple soul, he also shows contra Wolff and Crusius that even if the soul is simple this does not entail its immortality because it could perish through a remission of its powers.

Having argued that Kant's distinction between appearances and things in themselves may be employed to argue that the soul may possess multiple powers, I turn to a consideration of the role powers play in Kant's accounts of personhood, mind-body interaction, and freedom of the will. The third chapter (3) considers Kant's discussion of personhood in the Transcendental Deduction and the Third Paralogism. In contrast with Patricia Kitcher, who argues that the main historical context of Kant's discussion of personal identity is Hume's bundle account of the self, I argue that Kant's discussion is responding to the Lockean account of personal identity and the discussions of Locke's account of personal identity in Wolff and Leibniz.[19] The main concern in these discussions was with providing an account of personhood as a necessary condition for moral responsibility. Much like his empiricist and rationalist predecessors, Kant also develops a metaphysical conception of personhood that provides a necessary condition under which one can count as morally responsible for one's actions. According to Kant, personhood requires that we retain certain mental capacities – sensibility, understanding, and apperception – which make it possible for us to synthesize representations into continuous experience. And in contrast with a number of recent interpretations, I argue that Kant requires neither that representations be actually synthesized, nor that synthesis is accompanied by an awareness of the activity of synthesizing representations. I also argue in contrast with some recent interpretations that Kant believes that personhood would be retained even in cases in which consciousness or the capacities

[19] See Patricia Kitcher, "Kant on Self-Identity," *Philosophical Review* 91(1) (1982), pp. 41–72.

required for consciousness are transferred from one substance to another.[20] And I indicate that this account of personhood also supports a stronger understanding of a person in Kant as a moral agent by identifying the mental capacities needed for rationality and deliberation about moral choices.

The fourth chapter (4) considers Kant's critical solution to the problem of mind-body interaction in the conclusion to the Paralogisms of Pure Reason in the A and B editions of the *Critique of Pure Reason*. In his critique of Cartesian substance dualism, Kant argues that mind-body interaction is problematic for the dualist because the dualist has mistakenly taken mental and physical appearances to be things in themselves. Interpreters who favor an epistemological reading of Kant's distinction between appearances and things in themselves have argued that Kant merely means to argue against the dualist that the question whether and how mental and physical substances as things in themselves interact is unproblematic for us because we cannot know anything about things in themselves.[21] However, this interpretation is unsatisfying because it cannot explain Kant's appeals to noumenal affection, nor can it provide the resources for supporting Kant's account of the possibility of freedom of the will, which appears to require the possibility of mental causation. I argue instead that Kant provides a positive account of the possibility of mind-body interaction in the *Critique of Pure Reason*. Rather than positing a single power that allows for the interaction of substances as in *Thoughts on the True Estimation of Living Forces* (1747), Kant enlists transcendental idealism and the distinction between appearances and things in themselves to show how interaction is possible. According to Kant, it is possible that heterogeneous mental and physical appearances are grounded in things in themselves, which are intrinsically neither mental nor physical. Since there is no relevant difference in kind between the things in themselves that ground mental and physical appearances, their interaction is unproblematic. And I show how the explanation of interaction might be expanded by looking at Kant's discussion of the interaction of substances through their fundamental powers.

In chapter five (5), I consider Kant's critical solution to the problem of freedom of the will and determinism in the Third Antinomy. Recent metaphysical interpretations of Kant's discussion of the problem of freedom of the will and determinism such as those provided by Watkins and Vilhauer have argued that Kant's claim that freedom requires the "ability to

---

[20] See Julian Wuerth, "Kant's Immediatism–Pre-Critique," *Journal of the History of Philosophy* 44(4) (2006), pp. 489–532.
[21] See C. Thomas Powell, "Kant's Fourth Paralogism," *Philosophy and Phenomenological Research* 48(3) (1988), pp. 389–414

have done otherwise" entails that we have the ability to change the laws of nature.[22] I argue for an interpretation of Kant's compatibilist view of freedom of the will that does not require such an ability. For Kant, freedom of the will consists in the capacity to act from reason according to maxims. Although this capacity may be masked or undermined in certain circumstances by deterministic events, one nevertheless retains this capacity and so also one's freedom if and only if one retains the power of spontaneity. Because the retention of this capacity is independent of the ability to alter the laws of nature, Kant can provide a compatibilist account of freedom of the will without appealing to our ability to alter the laws of nature that govern empirical events. This interpretation also provides the foundation for answering a number of vexing interpretive and philosophical questions regarding the coherence of timeless agency, who may count as morally responsible, and what the extent of our moral responsibility is. It also reveals that Kant's conception of freedom as a capacity shares similarities with the views of Wolff, Tetens, and others. I conclude the dissertation by summing up the major arguments and pointing out some open questions regarding Kant's metaphysics of mind and its role in his theoretical and practical philosophy.

---

[22] See: Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005); Ben Vilhauer, "Incompatibilism and Ontological Priority in Kant's Theory of Free Will," in *Rethinking Kant: Volume I*, ed. Pablo Muchnik (Newcastle upon Tyne: Cambridge Scholars Publishing, 2008), pp. 22–47.

Chapter 1

Kant's First Paralogism and the Substantial Ground of Thought

1.1 Introduction

In the Transcendental Dialectic, Kant sets out to critique rationalist metaphysics regarding the soul, God, and the cosmos for illegitimately extending reason and the understanding beyond the bounds of experience. Kant criticizes rationalist views as dogmatic and unwarranted, but he also seeks to explain how his rationalist predecessors could be led quite naturally to their conclusions. They are not sophists according to Kant, but rather, they have succumbed to a "natural" and "inevitable" "illusion" that is due to the nature of human reason (A 298/ B 354). And the rationalist is led on the basis of this "transcendental illusion" to accept arguments that result in dogmatic conclusions about the soul, God, and the cosmos.[23] In the case of the Paralogisms, transcendental illusion leads the rationalist to think that whenever a conditioned thing is given, namely the attributes of thought, so too is its unconditioned ground, namely the soul. In the First Paralogism, Kant argues that the rationalist reasons that since the soul cannot be an attribute of anything else, it must be the unconditioned ground of the attributes of thought and therefore a substance. It is unclear, however, to many interpreters how the argument Kant attributes to the rationalist in the First Paralogism is supposed to function and whether the rationalist makes an a priori argument based on mere concepts or an a posteriori argument based on the putative observation of a substantial soul in inner sense or apperception in order to establish that the "I" or soul cannot be an attribute of another substance and why Kant believes the rationalist's argument involves "transcendental illusion."[24] How interpreters stand on these issues also informs whether they take Kant to be

---

[23] For an extensive treatment of the role of "transcendental illusion" in the *Critique of Pure Reason*, see Michelle, Grier, *Kant's Doctrine of Transcendental Illusion* (Cambridge: Cambridge University Press, 2001).

[24] There is a great deal of debate about whether the notion of transcendental illusion plays an explanatory role in Kant's critique of the paralogisms. Kitcher and Bennett reject the explanatory value of transcendental illusion for the paralogisms; see: Patricia Kitcher, "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), p. 518; Jonathan Bennett, *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974), p. 281. Grier and Proops argue for its explanatory value; see: Michelle Gilmore Grier, "Illusion and Fallacy in Kant's First

endorsing aspects of the rationalist's argument and the concluding idea that the soul is a substance. Some interpreters have argued, for example, that Kant attributes an a posteriori argument to the rationalist and that Kant in fact endorsed such an argument in the 1770's and the *Critique of Pure Reason*. On this view, Kant appears to endorse the idea that we can have an intuition of ourselves in apperception as the ground of thoughts with the caveat that this can reveal nothing about whether this ground of thoughts is immortal or not.[25] Others have taken Kant's aim to be primarily critical and maintain that he consistently rejects the rationalist's conclusion that the soul is a substance.[26]

One reason that Kant's discussion of transcendental illusion and the unconditioned ground of thought, his criticism of the rationalist's argument and conception of substance, and his apparent commitment to a substantial ground of thought remain so unclear is that interpreters have tended to overlook the historical context of Kant's criticism. Although Kant is clear that his criticism of the fallacies involved in the Transcendental Dialectic is not "a critique of books and systems, but a critique of the faculty of reason in general" (A xii), and so not aimed at any philosopher in particular, interpreters have often ignored the historical context of Kant's remarks or have uncritically taken Descartes to be the primary focus of Kant's discussion of rational psychology.[27] One exception in this regard has been Corey W. Dyck who has painstakingly traced the Wolffian tradition of rational psychology that Kant is

---

Paralogism," *Kant-Studien* 84(3) (1993), p. 258; Ian Proops, "Kant's First Paralogism," *Philosophical Review* 119(4) (2010), pp. 449–495.

[25] See: Julian Wuerth, "Kant's Immediatism–Pre-Critique," *Journal of the History of Philosophy* 44(4) (2006), pp. 489–532; Julian Wuerth, "The First Paralogism, its Origin, and its Evolution: Kant on How the Soul Both Is and Is Not a Substance," in *Cultivating Personhood: Kant and Asian Philosophy*, ed. Stephen R. Palmquist (Berlin: de Gruyter, 2010), pp. 157–166.

[26] For interpretations that focus on Kant's criticism of the rationalist and rejection of the substantial view of the self, see, for example: Ian Proops, "Kant's First Paralogism," *Philosophical Review* 119(4) (2010), pp. 449–495; Michelle Grier, "Illusion and Fallacy in Kant's First Paralogism," *Kant-Studien* 84(3) (1993), pp. 257–282.

[27] On Kant's criticism of Cartesian rational psychology, see Jonathan Bennett, *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974), pp. 66–81. Bennett takes Kant to be criticizing certain inferences made on the basis of the "Cartesian basis" of first-personal consciousness. Bennett interprets Kant as arguing that the self is not "in the world" but is rather "at the boundary of the world" and that therefore that knowledge of the self is elusive. The second edition Paralogisms provides no shortage of evidence for such a view. But as I will indicate, this reading overlooks the fact that Kant does not take the rationalist to be arguing from first-personal conscious experience for his claims but rather provides an a priori argument on the basis of the understanding of substance and the principle of sufficient reason.

criticizing and seeks to show that this tradition was always based on an admixture of elements of empirical psychology, which can be gathered from the observations of inner sense, and rational psychology, which is gained by a priori reflection.[28] Dyck argues that Kant sometimes focuses on a posteriori elements of the rationalist's argument in his criticism because the rational psychology Kant criticizes always contained such an admixture of elements of empirical and rational psychology. There is, however, a difference between the arguments for the nature of the soul provided by Wolff and the arguments provided by the German metaphysician Alexander Baumgarten whose *Metaphysica* (*Metaphysics*) (1739) Kant used throughout his career as the basis of his own lectures on metaphysics. Although a number of other rationalists could plausibly be construed as Kant's target in the discussion of transcendental illusion and the paralogistic argument for substantiality, Baumgarten is especially apt because, in contrast with the a posteriori arguments by Wolff and Baumgarten's student Georg Friedrich Meier, he provides an a priori argument for the substantiality of the soul that relies on the principle of sufficient reason and a conception of a substance as a persisting object. I argue that if we understand Kant as criticizing such a Baumgartian a priori argument for the substantiality of the soul, which is based in part on the principle of sufficient reason, then it becomes clearer how Kant's criticism of transcendental illusion is linked with his criticism of the rationalist's argument, which in turn opens the way toward understanding what aspects of the rationalist's conclusions regarding the self Kant may have endorsed and what his positive views on the substantial ground of thought may have been.

In my discussion of Kant's *Auseinandersetzung* with the rationalist argument for the substantiality of the soul, I proceed as follows. In section 1.2, I survey Baumgarten's a priori argument for the substantiality of the soul and argue that in the A edition Paralogisms Kant attributes to the rationalists an a priori argument for the existence of a substance as the ground of attributes that bears similarities to Baumgarten's argument. If we see Kant as criticizing Baumgarten's argument, then we are also able to make sense of Kant's claim that the reasoning in the first paralogism is motivated by transcendental illusion.[29] According to Kant, the rationalist mistakenly believes the application of the principle of sufficient reason

---

[28] See Corey W. Dyck, "The Divorce of Reason and Experience," *Journal of the History of the Philosophy*, 47(2) (2009), pp. 249–275.

[29] When referring to the Paralogisms as a chapter, I will capitalize the first letter, i.e. "Paralogisms." When referring to the particular paralogistic argument attributed to the rationalist, I will use "paralogism."

will yield an unconditioned ground of thought that is persisting substance. As Kant suggests, however, the unconditioned ground posited by the principle of sufficient reason cannot be an appearance but must be a thing in itself. Since it is a thing in itself, it cannot be a persisting substance, so the rationalist cannot conclude that such a substance is immortal. Kant allows, however, that the ground of thought may be a substance in some other sense. In section 1.3, I argue that Kant's lectures on metaphysics and passages from the *Critique of Pure Reason* suggest an account of how Kant may have understood the substance that underlies the attributes of thought. According to Kant, there is an unknowable unconditioned ground of thoughts or "substratum" that consists of intrinsic properties or powers that make the attributes of thought possible.[30] Section 1.4 concludes with a summary and raises questions that will be addressed in the next chapter. By looking more closely at the exact nature of the arguments for the substantiality of the soul that Kant may have had in mind in his discussion of transcendental illusion and the first paralogism, we not only gain an understanding of the historical provenance of Kant's discussion, but we also see stronger evidence of a continuity between Kant and his predecessors.[31]

1.2 The Rationalist Argument for a Substantial Ground of Thought

1.2.1 The Substantiality of the Soul

Although it is sometimes suggested that Kant's views on metaphysics in general were influenced to a great degree by rationalist philosophers such as Wolff and Baumgarten, there has been less attention paid to the precise nature of Kant's engagement with particular theses

---

[30] A number of passages suggest that Kant may have thought of the ground of the attributes of thought as a substance. He writes for example: "One can quite well allow the proposition The soul is substance to be valid, if only one admits that this concept of ours leads no further, that it cannot teach us any of the usual conclusions of the rationalistic doctrine of the soul" (A 351).

[31] Karl Ameriks argues in general that there is continuity between Kant's thought and the thought of his predecessors, but he does not engage with particular arguments provided by Wolff and Baumgarten. See Karl Ameriks, *Kant's Theory of Mind. An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000); Karl Ameriks, "The Critique of Metaphysics: Kant and Traditional Ontology," in *The Cambridge Companion to Kant*, ed. Paul Guyer (Cambridge: Cambridge University Press, 1992), p. 251.

and arguments provided by the rationalists, particularly regarding the nature of the soul.[32] This is perhaps because the rationalist positions are thought to be derived from Leibniz, and it is thought that nothing important can be found in the them that cannot be found already in Leibniz. Although it is true that Wolff and Baumgarten both owe a great deal to Leibniz, the German rationalist tradition has its own debates involving problems and distinctions, with which Kant was intimately familiar, that differ from those addressed by Leibniz. The subtlety of these debates is lost without looking directly at the exact nature of the arguments provided by these philosophers. As Corey W. Dyck has recently argued, the tradition of rational psychology that Kant is criticizing in his discussions of transcendental illusion and in the Paralogisms of Pure Reason was not in its time exclusively considered a pure a priori science divorced from the discoveries of empirical psychology. Rather, Wolff and others took empirical psychology to provide a foundation for the a priori aspects of rational psychology, which established the substantiality, simplicity, immortality, and freedom of the soul. To provide an example of this line of argument, Wolff argues that the deliverances of empirical psychology can be expanded through the use of a priori reasoning in rational psychology. So Wolff argues that we can perceive certain facts about the soul by attending to ourselves such as the fact that we are conscious of ourselves and of objects.[33] From the fact that we are conscious of ourselves and other objects, Wolff argues that we must have certain faculties that allow us to compare and distinguish among objects.[34] And he goes on to argue that the ability to compare and distinguish requires that "the soul is a simple thing" (§742), that it "subsists by itself" (§743), and that it has a single power from which its changes flow (§744–745).[35] Wolff concludes that since the soul is simple, it is incorruptible and therefore immortal. Importantly, Wolff's argument for the simplicity of the soul begins from the a

---

[32] For exceptions, see: Karl Ameriks, *Kant's Theory of Mind*. *An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000); Corey W. Dyck, "The Divorce of Reason and Experience," *Journal of the History of the Philosophy*, 47(2) (2009); Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005).

[33] See Christian Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (*Rational Thoughts on God, the World and the Soul of Human Beings, Also All Things in General*) (*Deutsche Metaphysik*) (1720) (Halle: 1751). On empirical psychology, see §191; on consciousness of ourselves, see §1.

[34] See Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (Halle: 1751), §729–734.

[35] See Christian Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (*Rational Thoughts on God, the World and the Soul of Human Beings, Also All Things in General*) (*Deutsche Metaphysik*) (1720) (Halle: 1751).

posteriori empirical fact that we are aware of ourselves and other things and builds upon this observation through a priori considerations about the nature of the substance, powers, and faculties required for such awareness. Thus the simplicity of the soul in Wolff's argument is not something that can be observed a posteriori or that is immediately known in inner sense, but it is a fact that can be established on the basis of empirical observation and reflection on ontological concepts.

Baumgarten may also be understood as incorporating a posteriori and a priori elements in his arguments regarding the soul in the *Metaphysica* (1739). He argues, for example, that the soul is a force of representing the universe according to the position of the body.[36] And he does so on the basis of an observation of the fact that we are capable of desire and repulsion.[37] However, although Baumgarten does present such arguments, there is nevertheless another route by which he argues for the substantiality of the soul, and the various properties that attach to substantiality such as immortality and freedom, on the basis of solely a priori considerations involving only reflection on ontological concepts such as "substance" and the principle of sufficient reason, which appears to be much closer to the kind of rational psychology, or "rational doctrine of the soul," that Kant criticizes in the Paralogisms and argues is "a putative science, which is built on the single proposition *I think*" that does not contain "the least bit of anything empirical" (A 342/ B400). Baumgarten's a priori line of argument for the simplicity of the soul or the "I" is quite distinct from the other lines of argument for the substantiality of the soul Kant would have been familiar with through either the Wolffians or other rationalists such as Descartes and Leibniz. Although it would be a difficult task to show decisively that Kant is thinking only of Baumgarten in his discussion of arguments for the simplicity of the soul, a close look at Baumgarten will help illuminate aspects of Kant's discussion in the First Paralogism that have previously remained obscure in a way that cannot be done if we take Kant's target to be some other rationalist philosopher with whom he was familiar.

---

[36] See Alexander Baumgarten, *Metaphysica* (Frankfurt: 1757), §505–13. Unless otherwise indicated, Baumgarten translations are from Eric Watkins (ed. and trans.), *Kant's Critique of Pure Reason: Background Source Materials* (Cambridge: Cambridge University Press, 2009).

[37] See Baumgarten, *Metaphysica*, § 741. See also Corey W. Dyck's discussion in "The Divorce of Reason and Experience," *Journal of the History of the Philosophy*, 47(2) (2009), p. 259.

According to Baumgarten, when we consider a thing, we see that it can exist either as a "determination of another" or it can exist otherwise than as a determination of another.[38] If it can exist only as the determination of another then it is "an accident," "whose being is in the being of another," and if it exists otherwise "then it is a substance […] which can exist, even if it does not exist in another, [that is] even if it is not a determination of another" (§191). Accidents or determinations are dependent beings in the sense that they could not exist if they were not grounded in a substance. Adopting a largely Aristotelian definition of substance, Baumgarten suggests that a substance is an independent being that cannot be an accident or determination of another.[39] Baumgarten reasons that all attributes must be grounded in such a substance by appealing to the principle of sufficient reason, or ground, (*Satz vom zureichenden Grund*, *principium rationis sufficientis*) according to which "everything possible has a sufficient ground."[40] He glosses this principle as holding that "having posited something, some sufficient ground of it is posited" (§22).[41] This is to say that if some determination exists, whether actually or merely possibly, then some sufficient ground for this determination also exists. A sufficient ground cannot, however, be another

---

[38] I use the terms "determination," "attribute," "accident," and "property" interchangeably.

[39] For a discussion of Baumgarten on substances, see Mario Casula, "A.G. Baumgarten entre G.W. Leibniz et Chr. Wolff," *Archives de Philosophie* 42 (1979), p. 564.

[40] See the introductory discussion of the principle of sufficient reason in Alexander Baumgarten, *Metaphysica*, ed. G. Gawlick and L. Kreimendahl (Stuttgart-Bad: Fromann-Holzboog Verlag, 2011), section 4.1. The German term '*Grund*' and the Latin term '*ratio*' mean both reason and cause. For other discussions of the notion of ground in the period, see also Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen* (*Deutsche Metaphysik*) (1720) (Hildesheim: George Olms, 1968), §29; Martin Knutzen, *Philosophische Abhandlung von der immateriellen Natur der Seele* (*Philosophical Treatise on the Immaterial Nature of the Soul*) (Königsberg: 1744), §5; Kant, Metaphysics L_2 (AA 28:572).

[41] Kant disputes Baumgarten's argument for the principle of sufficient reason in *New Elucidation of the First Principles of Metaphysical Cognition* (AA 1:397–398), and he expresses skepticism about any attempt to prove the principle of sufficient reason at A 783/B 811. For a discussion of Kant and the principle of sufficient reason, see Beatrice Longuenesse, "Kant's Deconstruction of the Principle of Sufficient Reason," *The Harvard Review of Philosophy* IX (2001), pp. 67–87. Leibniz also thought no argument could be provided for or against the principle of sufficient reason; see the Leibniz-Clarke correspondence, letters 2–5. For other attempts to argue for the principle of sufficient reason, see Wolff, *Deutsche Metaphysik*, §30, §31, and Georg Friedrich Meier, *Metaphysik*, Vol. 1 (Halle: 1755), §34. See also Christian August Crusius' critique of this principle in *Ausführliche Abhandlung von dem rechten Gebrauche und der Einschränkung des sogenannten Satzes vom zureichenen oder besser Determinierenden Grund* (Leipzig: 1744), §3, which was originally published in Latin as *Dissertatio philosophica de usu et limitibus principii rationis determinantis vulgo sufficientis* (Leipzig: 1743).

determination, because this determination would itself require another ground. So the sufficient ground of a determination must be an independent being, a substance. And all determinations must be grounded in such an "ultimate" ground, or "ground without qualification" (§28). Baumgarten also expands upon his definition of substance in his ontology by explaining the relation that must obtain between a substance and its determinations, or accidents, in order for an accident to inhere in a substance. He writes that "if accidents inhere in a substance, then there must be a ground of this inherence," and such a ground must be a "sufficient" ground. Moreover, such a ground is a "power" (§197). One might understand this to mean that substances possess a power that is a sufficient ground for the inherence of accidents, but Baumgarten maintains that a power is not itself an accident of a substance, for such an accident would itself require a sufficient ground, and so on this basis he is led to identify substances and powers. A substance grounds the inherence of its accidents by being a causal power that brings about the existence of its accidents. In this sense, a substance is not merely a bare substratum that exists after all determinations are removed but is also a sufficient causal and explanatory ground of its determinations.

Given this understanding of substances as powers that are the sufficient grounds of attributes, Baumgarten then applies these considerations to questions regarding the nature of the soul that grounds the attributes of thought. He argues that the soul cannot be a determination of another thing and so is an independent substance. This substantial soul is construed as an active ground of the inherence of the attributes of thought and the changes of thought. He writes: "Every spirit is a substance (§402), therefore, a power (§199), hence the sufficient ground of the inherence of its accidents (§197), and thus acting (§210)" (§755). After he has demonstrated that the attributes of thought must be grounded in a substance construed as a power, Baumgarten also argues in *Metaphysica* §756 that the soul has some additional properties that follow from its substantiality. He writes: "[T]he human soul is spirit (§754). Therefore, it has freedom (§755). And because spiritual, intellect, personality (§641, §754), freedom, absolute simplicity (§744), and incorruptibility are attributed to it with absolute necessity (§746)" (§756), and "the human soul is immaterial and incorporeal" (§757). Baumgarten derives these other properties of the soul from the fact that the soul is a substance. For example, since substances are simple, and therefore cannot be corrupted through the dissolution of their parts as a composite might be, the soul is incorruptible, i.e. immortal. Importantly, as becomes clear from Baumgarten's reasoning, he does not appeal to any facts established a posteriori on the basis of empirical psychology or the observation of thought or the I in inner sense as Wolff and others do. Rather, Baumgarten provides a wholly

a priori argument for the substantiality of the soul and its other properties by enlisting the principle of sufficient reason to show that the soul must be an ultimate substantial ground of the attributes of thought.

Kant's discussions of Baumgarten's rational psychology in his lectures on metaphysics suggest that he understood Baumgarten's conception of rational psychology and his arguments for the substantial ground of thought along precisely these lines. Kant is reported as saying in his lectures on metaphysics, which he delivered on the basis of Baumgarten's *Metaphysica*: "In rational psychology the human soul is cognized not from experience, as in empirical psychology, but *a priori from concepts*. Here we are to investigate *how much of the human soul we can cognize through reason*" (Metaphysik L_1, AA 28:262). Kant also continues that "when we consider the soul absolutely […], thus from transcendental concepts of ontology, then we will examine, e.g. whether the soul is substance or an accident" (AA 28:264).[42] This is to say that the examination of soul does not proceed through experience but through an a priori reflection involving other a priori concepts of rational ontology such as "substance," "accident," "power," and "sufficient ground."[43] Beginning from the concept of the soul, the rational psychologist considers whether it is a substance or an accident. Kant suggests that according to Baumgarten we can gather from a priori reflection on the soul "that it is a substance, or: I am substance." Expanding upon Baumgarten's reasons for concluding that the soul or I is substance, Kant says: "The *I* means the subject so far as it is not predicate of another thing. What is not predicate of another thing is substance. The I is *the general subject* of all predicates, of all thinking, of all actions, of all possible judgments that we can pass of ourselves as a thinking being. I can only say: I am, I think, I act. Thus it is not at all feasible that the I would be a predicate of something else. […] Consequently, the I, or the soul through which the I is expressed is a substance" (AA 28:266). As we have seen, Baumgarten does not offer any a posteriori reasons for thinking that the I cannot be a predicate of another thing, nor does Kant gloss his argument in this way. Rather, Kant notes that Baumgarten merely asserts that in all judgments the I serves as the subject of

---

[42] See also Kant, Metaphysik Mrongovius (AA 29:904f).

[43] In the discussion of empirical and rational psychology, Kant does sometimes appear to suggest both an a posteriori and an a priori route to establish the substantiality of the soul. The a posteriori route proceeds by arguing that we have an intuition of ourselves as a single unified thing and are therefore a substance. The a priori route argues that our representations must have a substantial ground and the soul must be this ground. See Heiner F. Klemme, *Kants Philosophie des Subjekts* (Hamburg: Felix Meiner Verlag, 1996), p. 113. However it appears that in the *Critique of Pure Reason* he focuses his criticism on the a priori argument.

attributes rather than as an attribute itself. The I also cannot be predicated of anything else. And since it cannot be predicated of anything else, it is a substance per Baumgarten's definition of substance as something that exists in such a way that cannot be a determination of another thing.

Although Kant is not explicit in his lectures about the method by which Baumgarten extends his conclusion that the soul is a substance rather than an attribute to broader conclusions about the nature of this substance, he is quite clear that rational psychology extends its conclusions on the basis of "transcendental concepts of ontology" (AA 28:264), which include concepts such as "power," "substance," and "sufficient grounds. Thus we see that Baumgarten argues that any substance that cannot be a determination of another thing must be a sufficient ground or power for the inherence of the attributes. As Kant notes elsewhere in his gloss of Baumgarten's discussion of powers: "Concerning power, it is to be noted: the author [Baumgarten] defines it as that which contains the ground of the inherence of accidents; since accidents inhere in each substance, he concludes that every substance is a power" (AA 29:771). Given his understanding of substance, Baumgarten also concludes that the soul is simple and immortal. It also appears that Kant understood the rationalist argument for the substantiality of the soul in the First Paralogism as an a priori argument very similar to that provided by Baumgarten and discussed by Kant at length throughout his lectures on metaphysics.

In the A edition First Paralogism, Kant attributes the following argument to the rationalist:

> That the representation of which is the absolute subject of our judgments, and hence cannot be used as the determination of another thing, is substance.
> I, as thinking being, am the absolute subject of all my possible judgments, and this representation of Myself cannot be used as the predicate of any other thing.
> Thus, I, as thinking being (soul), am substance. (A 348)

As it stands, there is some ambiguity in how Kant presents the argument, which makes it unclear whether the argument is supposed to be entirely a priori or whether it also involves an appeal to a premise that is established on the basis of a posteriori empirical observation. If "representation" means the ability to perceive and to represent the I as anything other than the subject of mental states on the basis of inner sense, then the argument appears to involve an appeal to the a posteriori deliverances of inner sense. This is to say that the argument hinges on whether in fact we can represent the I as anything other than a subject. It is, however, unlikely that Kant means this. For one thing, it would do a disservice to attribute such an argument to the rationalist. For even if it could be established that some particular person

may not be able to represent their I as anything other than the subject of mental states, there is no reason to think that such an empirical observation could be generalized for all subjects. Some subjects may be able to represent their I as a predicate of another thing and others may not. So it would be more charitable to the rationalist to maintain that the argument is intended to be entirely a priori. Understanding the argument as a priori also fits well with Kant's gloss on the argument.

Understood as an a priori argument, the major premise simply states a definition of a substance as something that cannot be represented in any other way than as a subject. And the minor premise claims that the I cannot be represented as a predicate but must be represented as a subject. And on the basis of these two a priori premises, it is concluded that the I, or soul, is a substance. Kant's subsequent discussion of the argument also makes the a priori character of the argument clear. He writes:

> Of any thing in general I can say that it is a substance, insofar as I distinguish it from mere predicates and determinations of things. Now in all our thinking the *I* is the subject, in which thoughts inhere only as determinations, and this I cannot be used as the determination of another thing. Thus everyone must necessarily regard Himself as a substance, but regard his thinking only as accidents of his existence and determinations of his state. (A 349)

As the argument is portrayed here, anything must be either a substance or a determination. Since the I cannot be thought of as or used as the determination of any other things, it must be a substance. And because the I must be thought of in this way, anyone with an I must regard themselves as a substance whose mental states are its determinations. Kant does not suggest that the rationalist appeals in any way to immediate consciousness of the I as a substance or to the empirical ability to represent the I as anything other than a subject of predicates. Rather, the argument suggests that the I cannot be conceived of as anything other than a subject. And it is the fact that the I cannot be conceived of as anything other than a subject that shows that it is a substance and that every thinker is such a substance. This I is also understood to be an "absolute subject" in the sense that there is no other subject or substance in which it could inhere. Kant is also quite explicit that the argument and its conclusion contain no a posteriori premises when he writes: "But now we have not grounded the present proposition on any experience, but have merely inferred [it] from the concept of the relation that all thought has to the I as the common subject in which it inheres" (A 350).

As it stands, the argument bears close similarities with the argument that Baumgarten makes in the *Metaphysica* and the argument Kant attributes to Baumgarten in his lectures on metaphysics. We have seen that in the lectures he interprets the Baumgartian argument as

saying: "it is not at all feasible that the I would be a predicate of something else. […] Consequently, the I, or the soul through which the I is expressed is a substance" (AA 28:266). In addition to the textual and historical evidence, there is also evidence by elimination. As we have seen, Wolff does not argue a priori for the substantiality of the soul because he sees a much closer connection between empirical and rational psychology. The same can be said of the other German philosophers such as Knutzen, Crusius, and Meier with whom Kant was familiar, each of whom appeal at least modestly to empirical observation in establishing their rational psychology.[44] Nor does it make sense to attribute such an argument to Descartes, who is often taken to be the target of Kant's discussion. In the Sixth Meditation, Descartes argues that the fact that we can clearly and distinctly conceive the mind or I as separate shows that it is an independent substance.[45] Although Kant may have discussed other arguments that attempt to demonstrate that the I is a substance on the basis of immediate consciousness of the I as substance or through some other means, such arguments do not appear to be the target of his discussion in the First Paralogism. Admittedly, such evidence is not incontrovertible, but it certainly strongly suggests that Kant's target in the First Paralogism was Baumgarten's a priori argument for the substantiality of the "I" or soul. More importantly, however, if we understand Kant as attributing an a priori argument to the rationalist similar to the argument provided by Baumgarten, then we are also better able to clarify the role Kant attributes to transcendental illusion in generating the rationalist's argument and to understand wherein Kant's critique of the a priori argument lies.

1.2.2 Transcendental Illusion

First we may consider transcendental illusion and then consider how Kant suggests transcendental illusion is involved in the rationalist's a priori argument for the substantiality of the I or soul. According to Kant, reason has the role of ordering concepts of the understanding into a systematic unity of thought, and it does this by following what Kant calls the "proper principle of reason in general (in its logical use)" (P1). This principle enjoins reason to "[F]ind the unconditioned for conditioned cognitions of the understanding, with which its unity will be completed" (A 307/B 364). This is to say that for any given conditioned thing, we should be led by reason to seek a more fundamental condition. For

---

[44] On the arguments provided by Crusius and Meier, see Corey W. Dyck, "The Divorce of Reason and Experience," *Journal of the History of the Philosophy*, 47(2) (2009).
[45] See René Descartes, *The Philosphical Writings of Descartes*, vol. II, trans. J. Cottingham, R. Stoothoff, D. Murdoch (Cambridge: Cambridge University Press, 1984), p. 54.

example, for some empirical occurrence E3, we always seek some cause E2 for the event. And for E2, we seek a further cause E1, and so on. This prescription of reason is not limited to the search for ever more fundamental causes but also applies to the search for the conditions for determinations that objects possess. Thus we seek for any determination some further ground of this determination. And we seek some further ground for conditioned things with the idea in mind that there is some unconditioned that grounds the various conditioned things, i.e. something that itself would not be subject to the search or further conditions.[46] The notion of an unconditioned ground is so to speak a goal toward which we strive.

Although Kant endorses this "proper principle of reason in general" as a guide to the systematic unity of our thought, he argues that we often mistakenly conflate it with a second principle, and in doing so we succumb to transcendental illusion. This second principle (P2) maintains that "when the conditioned is given, then so is the whole series of conditions subordinated one to another, which is itself unconditioned, also given" (A 307–8/B 364). What occurs when we succumb to transcendental illusion is that we take the prescriptive principle that enjoins us to find some further condition for conditioned things to be a metaphysical fact. And we think that whenever a conditioned thing is given, i.e. exists, so too is the final unconditioned. For example, reason demands on the basis of P1 that we seek for event E3 a cause E2, and for E2, we must in turn seek a cause E1. We succumb to transcendental illusion when we conflate P1 with P2 and thereby assume that when E3 is given so too is the entire series of events up to and including some unconditioned event EO. In this case, we posit the series of conditions as completed and given or existing in all cases in which the conditioned is given (A 309/ B 366). In this case "a logical prescription in the ascent to ever higher conditions to approach completeness in them and thus to bring the highest possible unity of reason into our cognition" has "through misunderstanding" "been taken for a transcendental principle of reason, which overhastily postulates such an unlimited completeness in the series of conditions in the objects themselves" (A 309/ B 366).[47]

It is illuminating to consider Kant's two principles in light of two principles that were often conflated by rationalists such as Baumgarten. In his discussion of grounds, or reasons,

---

[46] The application of the principle of reason in its legitimate use does not indicate if or where the search for conditions will terminate. In the *Prolegomena*, Kant entertains the possibility that the search for the grounds of thought may go on indefinitely; see AA 4:333f.

[47] For an extensive treatment of the role of "transcendental illusion" in the *Critique of Pure Reason*, see: Grier, Michele, *Kant's Doctrine of Transcendental Illusion*, N.Y.: Cambridge University Press, 2001.

Baumgarten introduces the "principle of reason" (PR), which holds that "nothing is without a ground," by which he means, for example, that for any event E3, there is always some ground of this event E2, and so on. This is to say that all things that exist have a ground. But this principle appears to leave open the possibility that this series of grounds could go in indefinitely. In fact, although Baumgarten does not say this, the principle that everything has a ground requires the series to go on indefinitely otherwise there would be some ground that itself does not have a ground, which would be in contradiction with the principle. Baumgarten, however, moves easily between the "principle of reason" and the "principle of sufficient reason [ground]" (PSR), which holds that "everything possible has a sufficient ground."[48] Whereas the first principle claims only that every event has some ground, the second principle claims that every event has some sufficient ground. The same is true of determinations and their grounds for Baumgarten. Whereas the principle of reason maintains that any determination has a ground, the principle of sufficient reason maintains that the series of grounds must terminate in a final sufficient ground.

Baumgarten's formulation of the principle of sufficient reason is very close to Kant's formulation of the illegitimate principle of reason P2. According to Baumgarten, the principle of sufficient reason maintains that "having posited something, some sufficient ground of it is posited." (§22). This is to say that if some determination is given, i.e. exists, then some sufficient ground of this determination also exists. Compare this with Kant's formulation of the illegitimate principle of reason (P2), which holds that "when the conditioned is given, then so is the whole series of conditions subordinated one to another, which is itself unconditioned, also given" (A 307–8/B 364). The similarity between Baumgarten's formulation of the principle of sufficient reason and Kant's formulation of the illegitimate principle of reason resides in their understanding of a final, absolute, or unconditioned ground. For Baumgarten, as we have seen, a sufficient ground is understood as a "first ground," beyond which no further ground can be given. Likewise, for Kant, an unconditioned ground of conditions is understood as an absolute condition for which no further condition

[48] Baumgarten and Wolff both conflate the principle of reason (*Satz vom Grund*; *principium rationis*) and the principle of sufficient reason (*Satz vom zureichenden Grund*; *principium rationis sufficientis*). Georg Friedrich Meier is more careful to distinguish between them in *Metaphysik*, where he says: "One must distinguish between the principle of reason and the principle of sufficient reason [*Grund*]. Namely, the reason [*Grund*] is either a sufficient or an insufficient reason [*Grund*]." See Meier, *Metaphysik*, vol. 1 (Halle: 1755), §34.

can be given.[49] The fact that Kant at least understood Baumgarten's formulation of the principle of sufficient reason along these lines can also be seen from Kant's gloss on Baumgarten's principle in the Metaphysik Mrongovius, where he is reported as saying: "The ground in the series of grounds subordinated to one another is only sufficient when we trace everything back to the first ground" (AA 29:817).[50]

What I would like to suggest is that Kant's legitimate principle of reason (P1) and the metaphysical principle (P2) can be understood along the lines of Baumgarten's principle of reason (PR) and the principle of sufficient reason (PSR). Kant appears to adopt something like Baumgarten's principle of reason (PR) and provides a legitimate use for it as a principle (P1) that guides our inquiry in the search for more fundamental grounds. Although Kant's P1 is not equivalent to Baumgarten's PR it does represent something like a revision of it. However, in contrast with Baumgarten who claims that PR entails PSR, Kant rejects the idea that P1 entails P2. This is to say that one may maintain the principle of reason as a guide to inquiry but reject the idea that there is some ultimate unconditioned ground that is given along with any conditioned thing. If we accept that P2 and PSR are equivalent, then we can see that transcendental illusion consists in conflating the legitimate principle of reason P1 with the principle of sufficient reason PSR. I also contend that the reason Kant believes these two principles should not be conflated is that the principle of reason P1 applies to appearances whereas the principle of sufficient reason (PSR or P2) does not. The principle of sufficient reason applies only to things in themselves and so cannot be applied to appearances. In this sense, although one may legitimately hold that any empirical appearance will have some further ground, this does not entail that any empirical appearances will have some absolute, i.e. sufficient ground. These issues are complicated and need to be spelled out in detail. We may now consider why Kant may have thought that the principle of sufficient reason does not apply to appearances, spell out in some detail what this might mean for Kant, and then consider the relevance of Kant's understanding of the principle of sufficient reason for his criticism of the a priori argument for the substantiality of I or soul.

---

[49] This understanding of the unconditioned as an absolute condition becomes especially clear in Kant's discussion of the thinking subject. See A 323/ B 379f; A 334/ B 391; A 348.

[50] Other interpreters have mentioned the proximity of the "supreme principle of pure reason" and the principle of sufficient reason but do not elaborate on the relationship; see: Henry Allison, *Kant's Transcendental Idealism. An Interpretation and Defense* (New Haven: Yale University Press, 1983), p. 53; Paul Franks, *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism* (Cambridge: Harvard University Press, 2005), p. 99.

One objection that might be raised immediately is that Kant appears explicitly to claim that the principle of sufficient reason applies to appearances. In the Second Analogy, Kant writes: "the principle of sufficient reason is the ground of possible experience, namely the objective cognition of appearances with regard to their relation in the successive series of time" (A 201/ B 246). This would seem to suggest that he thinks the principle of sufficient reason applies to appearances and that its application is necessary for experience. However, his explanation of the principle indicates otherwise. Kant explains the principle as maintaining that "that which follows or happens must succeed that which was contained in the previous state in accordance with a general rule" (A 200/ B 245). This is to say that for any event E2, there is some further event E1 such that E1 is the ground of E2 and E2 follows E1 in accordance with a general rule. And the application of this principle allows us to have coherent experience. But if we understand Kant's formulation of the principle of sufficient reason (P2, PSR) from the Transcendental Dialectic as maintaining that a conditioned event or thing has some unconditioned event or thing as its ground, then it is evident that this principle is not what is meant in the Second Analogy. In his explanation of the principle in the Second Analogy, Kant does not claim that we must regard an event as having some ultimate unconditioned ground that itself has no further ground; rather, we must simply regard any event as having some previous event that determines it according to a rule. Indeed, the fact that we must regard events as being determined by some previous event actually recalls Kant's formulation of the legitimate principle of reason (P1) rather than the principle of sufficient reason (P2, PSR). In other texts, Kant is more careful to call the principle referred to in the Second Analogy and the Transcendental Dialectic the principle of reason (*principium rationis*), which he allows only as a "rule of healthy reason" that is "restricted to the objects of experience" (R 4012, AA 17:385, 1769?).

But why should one think that the principle of sufficient reason in Kant's formulation of it from the Transcendental Dialectic does not apply to appearances? Why, for example, is it not legitimate to maintain that for any empirical event E2, there is some ultimate unconditioned empirical event E0 that has no further event as its ground? The problem appears to be that in positing an unconditioned ground for conditioned things, the principle of sufficient reason goes beyond anything that may be found in experience. Kant is clear about this when he writes, for example: "If they [concepts] contain the unconditioned, then they deal with something […] that is never itself an object of experience" (A 311/ B 367), and "the absolute totality of conditions is not a concept that is usable in an experience, because no experience is unconditioned" (A 326/ B 383). The principle of sufficient reason cannot apply

to appearances, i.e. to experience, because there are no unconditioned things in experience. The unconditioned ground of conditioned things cannot be an appearance. Why, however, does Kant think that "no experience is unconditioned" or that experience does not contain anything unconditioned? As we have seen, for Kant all experiences, i.e. appearances or empirical events, are conditioned. This is to say that we must regard any event E2 as following some previous event E1 according to a rule, and E1 must also be regarded as following some previous event according to a rule. This is a requirement expressed in the Second Analogy as necessary for experience. Now it appears contradictory to hold both that every event has some antecedent determining event and that there is some event that is an ultimate or first event that does not have some antecedently determining event. Thus it cannot be that both the principle of reason (P1) and the principle of sufficient reason (P2, PSR) apply to appearances. And we have seen that Kant is adamant that the former does apply to experiences. So the latter does not.[51] It also appears that the category of relation, which is in part constitutive of the unity of experience, does not allow for the possibility that in the relations – of inherence and subsistence, causality and dependence, community – there could be some absolute unconditioned *relatum* that is not itself subject to such relations.[52] Now if no experience can be unconditioned, then it appears that the principle of sufficient reason, which makes reference to the unconditioned ground of some conditioned thing, could not apply to the world of appearances. Kant is also explicit about this when he criticizes Eberhard, a defender of the principle of sufficient reason, in *On a Discovery* (1790), arguing that the principle of sufficient reason extends beyond our capacity for knowledge (AA 8:198). It extends beyond our capacity for knowledge in the sense that it extends to unconditioned conditions, which are not themselves within the domain of the appearances or objects of experience of which we can have knowledge.

I suggest that Kant warns that we should not conflate the principle of reason (P1), which has a legitimate application to appearances, and the principle of sufficient reason (P2, PSR), which does not legitimately apply to appearances. However, although Kant rejects the idea that the principle of sufficient reason has a legitimate application to appearances, this

---

[51] Kant's claim in the Second Paralogism that "it is also impossible to derive [the] necessary unity of the subject, as a condition of the possibility of every thought, from experience. For experience gives us cognition of no necessity, to say nothing of the fact that the concept of absolute unity is far above its sphere" (A 353) also appears to suggest that an absolute ground could not be an object of experience.

[52] See A 80/ B 106.

does not entail that it has no application at all. Although Kant does not say this explicitly, it appears that the only domain of application for the principle of sufficient reason would be to things in themselves. Although all appearances must be subject to further conditions, things in themselves need not be. This is evident, for example, in Kant's discussion of freedom of the will. Here he argues that we are causally determined as appearances, i.e. our actions are such that there is always some antecedently determining action that determines our subsequent actions according to a rule, but as things in themselves our actions are not antecedently determined, i.e. that we have the ability to begin an action freely without being subject to antecedent conditions.[53] It appears that any unconditioned ground would have to exist outside of appearances. And it appears that the only possibility here is that an unconditioned ground is a thing in itself if it is anything at all. So in the search for an unconditioned condition posited by the principle of sufficient reason, we must look beyond appearances to things in themselves. We may now consider how Kant's understanding of the principle of sufficient reason and transcendental illusion inform his criticism of the rationalist's a priori argument for the substantiality of the soul.

1.2.3 The Rationalist's Error

If it is true that Kant maintains that the principle of sufficient cannot legitimately be applied to experiences because it posits an unconditioned condition that could only be a thing in itself, then we are closer to understanding why Kant maintains that the a priori rationalist argument for the substantiality of the I or soul is guilty of the transcendental illusion of conflating the principle of reason (P1) and the principle of sufficient reason (P2, PSR).[54] Referring to the arguments provided by the rationalist regarding the soul, the cosmos, and God, Kant writes: "there will be syllogisms containing no empirical premises, by means of which we can infer from something with which we are acquainted to something of which we have no concept, and yet to which we nevertheless, by an unavoidable illusion, give objective

---

[53] For such an interpretation of Kant's views on freedom of the will, see Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, ed. Allen Wood (Ithaca: Cornell University Press, 1984), pp. 73–101.

[54] Neither Proops nor Grier see the connection between Kant's critique of transcendental illusion and the principle of sufficient reason in their discussion of the role of transcendental illusion in the Paralogisms. Omri Boehm has recently argued for such a connection but does not consider its relevance for the Paralogisms; see Omri Boehm, "The Principle of Sufficient Reason, the Ontological Argument and the Is/Ought Distinction," *The European Journal of Philosophy* (forthcoming).

reality" (A 339/ B 397). According to Kant, the rationalist in his a priori argument carries out a prosyllogism employing the principle of sufficient reason to "proceed to the unconditioned." Kant does not provide a detailed description of how this occurs, but he appears to mean that in the case of the first paralogism, the rationalist reasons as follows. The rationalist is acquainted with thoughts. On the basis of the acquaintance with such thoughts, the rationalist argues that these thoughts must have some ground. The rationalist argues in this way because he applies the principle of reason (P1) and thus seeks a condition for every conditioned thing. The rationalist holds that the I must be such a condition for thoughts. The rationalist is, however, guilty of transcendental illusion, when he argues that the I is an "absolute subject." This is to say that the rationalist maintains that the I is a ground or condition beyond which there can be no further ground or condition. The illusion occurs because the rationalist thinks that whenever some conditioned thing, in this case thought, is given, so too is the unconditioned condition of this thing. Moreover, the rationalist is then led to give this I as absolute subject "objective reality," by which Kant appears to mean that the rationalist maintains that such an absolute subject is a possible object of experience. Problematically, however, it is unclear whether the formal argument Kant attributes to the rationalist in the First Paralogism actually proceeds in this way. Kant leaves far too much unexplained in his discussion to see decisively that this is the line of reasoning he attributes to the rationalist.

In his description of the rationalist's argument for the substantiality of the soul in the *Prolegomena*, Kant is only slightly clearer about why the rationalist's argument involves transcendental illusion. He writes:

> Pure reason demands that for each predicate of a thing we should seek its appropriate subject, but that for this subject, which is in turn necessarily only a predicate, we should seek its subject again, and so forth to infinity (or as far as we get). (AA 4:333)

As we have seen, the principle of reason (P1) demands that we seek a condition for all conditioned things. In the *Prolegomena* discussion of the rationalist argument, this means that the rationalist searches for a more fundamental subject for every predicate. Regarding the rationalist's reasoning, Kant writes:

> Now it does appear as if we have something substantial in the consciousness of ourselves (i.e., in the thinking subject), and indeed have it in an immediate intuition; for all the predicates of inner sense are referred to the *I* as subject, and this *I* cannot again be thought as the predicate of some other subject. It therefore appears that in this case completeness in the referring of the given concepts to a subject as predicates is not a mere idea, but that the object, namely, the *absolute subject* itself, is given in experience. But this expectation is disappointed. (AA 4:334)

In our pursuit for a condition for conditioned things, we appear to encounter an absolute subject in consciousness. All the predicates of inner sense are attributed to the I, and the I itself cannot be thought of as the predicate of some other subject. In this regard, the I is taken to be an absolute subject. As Kant says, "in this case completeness in the referring of the given concepts to a subject as predicates is not a mere idea." This is to say that the notion of an unconditioned condition is not something that reason pursues as a mere idea in searching for ever more fundamental conditions. Rather, "the object, namely, the *absolute subject* itself, is given in experience." The unconditioned condition of thought appears to be given in experience as the absolute subject of thoughts. It appears that the rationalist begins with the legitimate principle of reason (P1) but conflates this principle with the principle of sufficient reason (P2, PSR), which leads the rationalist to think that he has found the unconditioned condition posited by the principle of sufficient reason in experience. Although Kant does not say this, one might surmise that the rationalist is led to believe that the unconditioned condition of thought is something that can be encountered in experience because the rationalist fails to distinguish between appearances and things in themselves and thus takes the principle of sufficient reason to apply to appearances.

Although Kant is notoriously unclear about the role transcendental illusion plays in the rationalist's conclusions regarding the substantiality of the soul, we are now in a better position to understand his criticism of the rationalist. Because the rationalist conflates the search for an unconditioned ground with the idea that such an unconditioned ground is to be encountered in experience, the rationalist also draws false conclusions regarding the substantiality of the I or soul. And the false conclusions regarding the substantiality of the soul also lead the rationalist to further false conclusions about the persistence and immortality of the soul. As we have seen, for Kant, if there is an absolute unconditioned ground of thought, it will be something that is not an appearance but a thing in itself if it is anything at all. In his criticism in the First Paralogism of the rationalist's conclusion that the soul is a substance, Kant argues that "pure categories (and among them also the category of substance) have in themselves no objective significance at all unless an intuition is subsumed under them" (A 348–9). This means that in order for the rationalist to apply the concept of a substance as "the persistence of the real in time" (A 144/B 183) to the I or soul it must have an intuition of this I in experience. But as Kant notes, the rationalist has no "standing and

abiding" intuition of the I.[55] So the I cannot be regarded as a substance in the sense of something that persists across time. And if the I cannot be regarded as a substance in this sense, then the rationalist cannot conclude that the I or soul is immortal. This criticism of the rationalist is well known. However, given the preceding discussion we see that there is more to Kant's story. In the search for the ground of thoughts, the rationalist posits that the soul is the absolute subject of thought, i.e. that it is a subject that itself cannot be a determination of another thing. But as we have seen, Kant notes that an absolute subject could not in principle be an object of possible experience. And since such an absolute subject could not be an object of possible experience, the concept of a substance as a persisting thing cannot be applied to it. The problem is that the rationalist cannot claim both that the I is the absolute unconditioned subject of thought and that it is a persisting substance. If it is an absolute subject, it cannot be a persisting substance, and if it is a persisting substance, then it cannot be an absolute subject. The rationalist is guilty of applying a spatial and temporal conception of substance to something that could only be a thing in itself.

We began the discussion by suggesting that a close look at Baumgarten's argument for the substantiality of the soul will put us in a position to illuminate aspects of Kant's criticism of the rationalist argument for the substantiality of the soul. As we have seen, Baumgarten reasons just as Kant suggests the rational psychologist reasons. He begins from the attributes of thought and searches for a ground for these thoughts. Because he conflates the principle of reason (PR) with the principle of sufficient reason (PSR), he believes that whenever a grounded thing is given so too is its sufficient or final ground. Since the I cannot be conceived of as an attribute but only as a ground, he concludes that it is a final or sufficient ground of the attributes of thought. And because of his understanding of the nature of sufficient grounds, Baumgarten concludes that the I must be a persisting substance and that it is therefore immortal. However, because Baumgarten does not distinguish between appearances and things in themselves, he takes the principle of sufficient reason to apply to all things. Thus he believes that the absolute subject of experience is something that can be given in experience. But as Kant suggests, the absolute subject posited by the principle of

---

[55] In the mid 1770s, Kant appears to have thought that we could have an immediate intuition of the substantiality of the soul, but it is unclear whether he endorses or merely considers this claim. See *Metaphysik* L_1 (AA 28:226; AA 28:266). For a thorough discussion of Kant's objection to the rationalist's use the schematized category of substance in the Paralogisms, see Dina Emundts, "Die Paralogismen und die Widerlegung des Idealismus in Kants *Kritik der reinen Vernunft*," *Deutsche Zeitschrift für Philosophie* 54(2) (2006), pp. 295–309.

sufficient reason must be a thing in itself if it is anything at all. And since it is a thing in itself, the concept of a persisting substance cannot possibly apply to it, and thus it cannot be considered immortal.

In the preceding discussion, I have argued that Kant maintains against rationalists such as Baumgarten that the application of the principle of sufficient reason regarding the absolute subject of thought cannot yield a subject of thought that is a persisting substance. But this is not to say that Kant rejects entirely the idea that there may be a substance that is the unconditioned ground of the attributes of thought. Since, however, such an unconditioned ground or absolute subject would have to be a thing in itself if it is anything at all, a different notion of substance must apply to it than that of a persisting substance. I turn now to a consideration of Kant's thoughts on the ground of thought in order to show that in the First Paralogism and in his lectures on metaphysics Kant appears to be sympathetic to the idea the ground of thought is a substance in some special restricted sense.

1.3 Kant's Conception of the Substantial Ground of Thought

Despite his rejection of the rationalist's claim that the absolute ground of the attributes of thought is an empirical substance, Kant nevertheless maintains that there is a "substratum" that grounds thoughts although "we have no acquaintance with this subject in itself that grounds this I as a substratum, just as it grounds all thoughts" (A 350). His recognition that there is such an unconditioned ground of thought also leads Kant to suggest that "one can quite well allow the proposition *The soul is substance* to be valid, if only one admits that this concept of ours leads no further, that it cannot teach us any of the usual conclusions of the rationalistic doctrine of the soul […]" (A 350–1). The ultimate ground of thought is a substance as Kant notes, but it is not a substance in the sense that the rationalist requires in order to argue that the soul is immortal. Since the spatial and temporal conception of a substance does not apply to the ultimate unconditioned ground of the attributes of thought, the rationalist cannot conclude anything about the persistence or immortality of this ground. Nevertheless Kant does suggest that some restricted sense of substance does apply to the ground of thought "as it is in itself".[56] Kant's suggestion that the ground of thought is a

[56] The fact that Kant retained some commitment to a substantial ground of the attributes of thought is perhaps not entirely surprising given his commitment to this view in some form or

substance is not isolated to the *Critique of Pure Reason* but can also be found throughout the *Duisburg Nachlaß* and the lectures on metaphysics.[57] In a *Reflexion* from the period 1780–1789, he writes for example: "The soul in transcendental apperception is *substantia Noumenon* [noumenal substance]; therefore no permanence of the same in time; and this can be only for objects in space" (R 6001, AA 18:420–1). And in a later *Reflexion* from 1795, Kant also opposes the schematized conception of substance with an acceptable conception of a substance that may characterize the ground of the attributes of thought, writing: "It appears that, if one admits that the soul is substance, one also needs to admit permanence as with bodies. But we can recognize absolutely nothing permanent in the soul, as, e.g., heaviness or impenetrability with bodies. – Thus the concept of the soul as substance is only a concept of a bare category of the subject to distinguish it from its inhering accidents" (R 6334, AA 18:655).[58]

Kant often contrasts the spatial and temporal concept of substance as a persisting thing with the pure concept of substance as something that cannot be thought otherwise than as subject.[59] Such a distinction can be seen in the preceding *Reflexion* and in "On the Schematism of the Pure Concepts of the Understanding," where Kant writes: "if one leaves out the sensible determination of persistence, substance would signify nothing more than a something that can be thought as a subject (without being the predicate of something else)" (A 147/ B 187).[60] One commentator has recently argued that the concept of substance that Kant believes may appropriately be applied to the thing in itself that grounds the attributes of thought is the concept of an "indeterminate" substratum distinct from all its attributes.[61] The

---

another throughout his pre-critical writings. For a discussion of Kant's pre-critical views on the soul as substance, see Alison Laywine, *Kant's Early Metaphysics and the Origins of the Critical Philosophy* (Ridgeview Publishing Company, 1993).

[57] See Wolfgang Carl, *Der schweigende Kant* (Göttingen: Vandenhoeck & Ruprecht, 1989), pp. 91–3. See also R 4684, AA 17:672. Here Kant understands the soul as a simple, immaterial substance.

[58] Translations of R 6001 and R 6334 are my own.

[59] For additional discussions of Kant's views on substance, see: Heinz Heimsoeth, *Studien zur Philosophie Immanuel Kants*: *Metaphysische Ursprünge und Ontologische Grundlagen* (Köln: Kölner Universitäts Verlag, 1956), pp. 75, 149, 194; Peter Schulthess, *Relation und Funktion: Eine systematische entwicklungsgeschichtliche Untersuchung zur theoretischen Philosohie Kants* (Berlin: de Gruyter 1981), p. 157.

[60] See also R 5295, AA 18:145.

[61] See: Julian Wuerth, "The First Paralogism, its Origin, and its Evolution: Kant on How the Soul Both Is and Is Not a Substance," in *Cultivating Personhood in Kant and Asian Philosophy* (New York: De Gruyter, 2010), pp.157–166, p. 164; Julian Wuerth, "Kant's

temptation to call this ground indeterminate is understandable given Kant's claim that the substratum is what remains after all determinations are removed. But there are some reasons also to doubt that the concept of substance Kant believes applies to the ground of the attributes of thought is the concept of a wholly indeterminate substratum. Kant is quite clear, for example, that spatial and temporal properties do not apply to things in themselves and so do not apply to the thing in itself or things in themselves that ground the attributes of thought. If this is the case, then it is determinate that the thing in itself that grounds the attributes of thought is neither spatial nor temporal. Likewise, in the Second Paralogism Kant denies that the property of material compositeness applies to things in themselves because such a property is based in our spatial and temporal forms of intuition and inner and outer sense. If this is the case, then it is determinate that the unknowable substratum of empirical thoughts is not composite in the way that physical bodies are composite. More importantly, however, there are some worrisome implications for Kant's philosophy if the substratum that grounds thoughts is taken to lack all determinations. For example, it is unclear how one substantial ground of thought is individuated and distinguished from another substantial ground of thought if they lack determinations whereby they could be distinguished. It is important, however, that things in themselves are individuated in some way since Kant eventually argues in the Third Antinomy that the things in themselves that ground empirical thoughts and actions are the source of moral responsibility for persons. If things in themselves were not individuated, then the moral responsibility for an action would apply to a single noumenal ground rather than individual noumenal persons. Part of the motivation for the idea that the substratum that grounds appearances must be indeterminate appears to be driven by epistemological concerns. It is thought that since determinations are made on the basis of judgments and knowledge, something of which we can have no knowledge has no determinations. But there is no reason to think that Kant would have endorsed this epistemic restriction. Although the thing in itself that grounds thought may lack the determinations that can be attributed to it on the basis of knowledge claims, this does not mean that it lacks even mind-independent or knowledge-independent determinations. Moreover, Kant also appears to suggest that things in themselves are completely determinate.[62]

Immediatism–Pre-Critique," *Journal of the History of Philosophy*, 44(4) (2006), pp. 489–532.

[62] On the issue of whether Kant held that things in themselves are completely determinate, see: Lucy Allais, "Kant's Transcendental Idealism and Contemporary Anti-realism," *International Journal of Philosophical Studies* 11(4) (2003), pp. 369–392; James Van Cleve,

This is not to say that the above interpretation of Kant's views on noumenal substances as lacking determinations is without any merit. Part of the confusion in interpreting what conception of substance Kant might have believed applies to the ground of the attributes of thoughts and why he might think it applies can be attributed to the fact that Kant often vacillates in his discussions of substances between a Lockean conception of a substance as an indeterminate substratum that exists after all determinations have been removed and a Baumgartian conception of a substance as a power. In a passage from Metaphysik Mrongovius, Kant offers some reasons for rejecting the Baumgartian conception of substance in favor of a Lockean conception:

> Concerning power, it is to be noted: the author [Baumgarten] defines it as that which contains the ground of the inherence of accidents; since accidents inhere in each substance, he concludes that every substance is a power. This is contrary to the rules of usage: I do not say that substance is a power, but rather that it has a power, power is the relation <respectus> of the substance to the accidents, insofar as it contains the ground of their actuality […]. We have absolutely no acquaintance with the substantial, i.e. the subject, in which no accidents inhere, which must be necessarily distinguished from the accident, for if I cancel all positive predicates then I have no predicates and cannot think anything at all. (AA 29:771)

Baumgarten conceives of a power as the ground for the inherence of accidents. Since, as we have seen, all accidents must have a sufficient ground in a substance, the power that grounds accidents must itself be a substance. Kant, however, warns his students that it is a linguistic, or worse, a category mistake to call a substance a power.[63] This is because a power is the relation of a substance to an accident, i.e. it is that which allows a substance to ground its accidents. As such, a power is itself an accident of a substance rather than a substance itself. Having dismissed Baumgarten's claim that a substance is a power, Kant then returns to the Lockean conception of a substance as a bare substratum and suggests that since a power is an accident it must have its ground in a bare substratum.[64] However, although Kant appears to

---

Problems from Kant (New York: Oxford University Press, 1999), pp. 224f. See also Kant's discussion of the principle of complete determination at A 571f./ B 599f.

[63] For discussions of substances and causal powers in Kant, see: Stefan Heßbrüggen-Walter, *Die Seele und ihre Vermögen: Kants Metaphysik des Mentalen in der Kritik der reinen Vernuft* (Paderborn: Mentis Verlag, 2004); Andree Hahmann, *Kritische Metaphysik der Substanz: Kant im Widerspruch zu Leibniz* (Berlin: Walter de Gruyter, 2009). Kant also discusses powers, substances, and accidents, in Metaphysik L_2 (Pölitz) (1790–1791?), AA 28: 564f.

[64] It is an open interpretive question whether Locke actually endorses the idea of substance as a substratum as "a supposed, I know not what, to support those *Ideas* we call Accidents." See John Locke, Essay Concerning Human Understanding, ed. Peter H. Nidditch (Oxford:

argue unequivocally against Baumgarten that a substance must be considered a bare substratum rather than a power, he also suggests that power is "something that is not substance, yet also not accident" (AA 29:771), which suggests some confusion on his part about the topic.[65]

Kant's confusion can be clarified, however, if we recognize that Kant saw that something more was needed than a Lockean substratum conception of substances in order to explain how substances ground their attributes. Kant is quite clear that he agrees with Baumgarten that a power is the particular kind of accident that allows for the inherence of other accidents in a substance. But if only a power can ground attributes, it is unclear how a bare substratum can be the ground of attributes. Such a substratum is only that which remains when one abstracts all attributes including those attributes that Kant calls powers. Such an entity is, however, metaphysically inert in the sense that it does not have what it takes to ground attributes. So Kant must agree with Baumgarten that only a power can be the ultimate or absolute ground of attributes. But he need not agree that such a power is to be equated with a substance. One way to look at this it suggest that for Kant the powers of a substance are the intrinsic properties of a substance that determine what kind of substance a substance is. These intrinsic properties of a substance are what individuate the substance as a particular individual substance. And they are also the means whereby the substance grounds its various other extrinsic attributes.[66] This might be illustrated with the example of the power of thought. The power of thought is an intrinsic property of a substance, which along with other intrinsic properties individuates the substance. This power of thought also provides the ground for various extrinsic attributes, such as particular thoughts that arise in conjunction with sensibility. So, for example, the thought "it is snowing" is an extrinsic attribute of a

---

Clarendon Press, 1974), II.xxiii.15; see also II.xxiii.2. Locke often discusses substances alternately as a bare logical substratum and as a real essence or structure underlying qualities. For a defense of the former as Locke's view of substance, see Jonathan Bennett, *Locke, Berkeley, Hume* (Oxford University Press, Oxford, 1971) and Bennett, "Substratum," *History of Philosophy Quarterly* 4 (1987), pp. 197–215. For a defense of the latter, see Michael Ayers, "The Ideas of Power and Substance in Locke's Philosophy," *Philosophical Quarterly* 25 (1975), pp.1–27, and Ayers, *Locke*, vol.2, part I (London: Routledge, 1991), pp. 26–42. In favor of the substratum conception, Kant also says: "If we leave aside all accidents then substance remains, this is the pure subject in which everything inheres or the substantial, e.g., I. All powers are set aside here" (AA 29:771).

[65] Longuenesse also discusses Kant's notion of substance within the context of German rationalism; see Beatrice Longuenesse, *Kant and the Capacity to Judge* (Princeton: Princeton University Press, 1998), p. 332.

[66] On the notion of a real ground, see R 4412, AA 17:536–37 (1771).

substance, which is grounded in the power of thought, which is an intrinsic attribute of the substance. Although one might logically abstract the intrinsic attributes of a substance away in order to arrive at the notion of a bare substratum, such a substratum is not a metaphysically basic entity since the substance cannot exist without its intrinsic properties, i.e. its powers. By interpreting Kant this way, we see he is able to navigate a middle way between the Lockean and Baumgartean conceptions of substance by agreeing that there is indeed an abstract logical notion of a substance as a bare substratum but that such a substratum is not sufficient to ground the attributes of thought since it does not possess the powers that would allow it to do so. And he is able to concede to Baumgarten that only a power can ground attributes while maintaining that such a power is not itself identical to a substance but is an intrinsic property of a substance without which a substance could not be the substance that it is.

The suggestion that for Kant a power is an intrinsic property of a substance whereby it grounds its extrinsic attributes is not, however, impervious to objection. Rae Langton has argued that intrinsic properties cannot be causal powers for Kant. She provides a simple thought experiment to illustrate this point. We can imagine two worlds that are identical in terms of the intrinsic properties they contain but differ according to the natural laws that obtain in the respective worlds. This suggests that the causal powers that determine the natural laws are extrinsic properties rather than intrinsic properties. Langton develops this position by arguing that Kant held that causal powers are superadded to objects. On this view, God chooses a set of objects with their intrinsic natures, but this choice of objects does not determine the causal powers that these objects have. "God superadds (*insuper accesserit*) to the monads powers of relating to each other, and this creative act is entirely 'arbitrary', and can be omitted or not omitted at God's pleasure."[67] This contrasts, for example, with God's choice of two individuals with their respective intrinsic properties and their relative sizes. God's choice of these individuals and their properties determines what relational properties these individuals will have in terms of their relative sizes. The fact that one will be taller than the other, or shorter than the other, or that their heights will be equal is determined by God's choice of individuals. In this case, relational properties of size supervene on the intrinsic

---

[67] See Rae Langton, *Kantian Humility: Our Ignorance of Things in Themselves* (Oxford: Oxford University Press, 1998), p. 118. For a detailed discussion of Langton's views on causal powers in Kant, see Lucy Allais, "Intrinsic Natures: A Critique of Langton on Kant," *Philosophy and Phenomenological Research* 73(1) (2006), pp. 143–169. Allais argues against Langton's proposal.

properties of the individuals. God's choice of individuals with their intrinsic properties does not, however, determine what causal powers will obtain, so causal powers do not supervene on the intrinsic properties of the individuals.

But there are several reasons why the proposed interpretation of Kant's conception of substance would not be subject to such an objection. First, there is little direct textual evidence in the *Critique of Pure Reason* that Kant thought that causal powers are superadded. And it also appears to follow from Kant's view on the freedom of the noumenal self that two identical noumenal substances would in fact instantiate the same natural laws. This is because the natural laws that exist are due to the intrinsic nature, or character, of things as they are in themselves.[68] Moreover, as I have argued, we can see from Kant's confrontation with Baumgarten's conception of substance that there are strong reasons for thinking that Kant may have thought that powers are in fact intrinsic properties of substances and that it is only in virtue of these powers that a substance can ground its attributes. If certain powers are not identical with intrinsic properties but must themselves be grounded in intrinsic properties, then it is unclear how intrinsic properties ground such powers without at the same being a power. If only a power can ground an attribute including an essential attribute such as an intrinsic property, then intrinsic properties that ground powers must themselves be powers, which is just what has been argued.[69] Although it is true that the fact that the proposed interpretation of Kant's view on substances is not proven by the fact that it can handle the kind of objection proposed by Langton, one obstacle to accepting the interpretation has at least been removed.

Another objection that might be raised against interpreting Kant as holding that there are things in themselves that ground the attributes of thought through their powers is that it would be inconsistent for Kant to maintain this view and claim that we cannot have knowledge of things in themselves. If we cannot have knowledge of things as they are in themselves, then how could Kant posit that there are things in themselves endowed with certain powers? This kind of objection is an instance of a more general type of objection

---

[68] Eric Watkins argues that the laws of nature are ultimately determined by the nature of things in themselves. See Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005), p.334.

[69] It is important, however, to recognize that the notion of a power that Kant endorses is not that of a phenomenal causal power. If he were talking about a phenomenal cause, he would be guilty of making a mistake analogous to the one he accuses the rationalist of making regarding the illicit use of the phenomenal concept of substance. For Kant, a power is simply that whereby an object grounds its attributes.

regarding Kant's account of things in themselves. In his account of things in themselves, Kant appears to claim both that they are the grounds of appearances and that we cannot have knowledge of them. But as has often been pointed out, if we cannot have knowledge of things in themselves, Kant cannot legitimately claim that they ground appearances. And so much less so would he be able to say that things in themselves ground appearances through their powers or that things in themselves ground the attributes of thought through their powers. It would be well beyond the scope of this paper to suggest that I can provide a decisive argument regarding how Kant's seemingly contradictory claims can be reconciled, but there are few things that might be said here. First, it is widely accepted that there is ample textual evidence that Kant claims that we cannot have knowledge of things in themselves and nevertheless ascribes certain positive attributes to them.[70] The claims regarding the things in themselves that ground thought may simply be another instance of this tendency in Kant. One might simply bite the bullet here and accept that Kant is in fact inconsistent about his claims. If this is true, then one has the option of choosing the version of Kant one minds most palatable, the one committed to the idea that things in themselves ground appearances or the one who wholly rejects any substantive claims about the nature of things in themselves. It has also often been argued that the most charitable way to interpret Kant's seeming confusion is to suggest that in the A edition, Kant's commitments to claims about things in themselves are very present, but that he attempted to eliminate the inconsistency in the B edition.[71]

However, I think there is more promise in showing that Kant's claims about our ignorance of things in themselves are not in fact incompatible with his positive claims regarding things in themselves. Let's grant that Kant says that we cannot have knowledge (*Erkenntnis*) of things in themselves. One way to understand this claim is in a weak sense as meaning that we cannot cognize things in themselves. For Kant, cognition requires judgments, which require both intuitions and concepts.[72] It is true that Kant maintains that we cannot cognize things in themselves in this sense. We cannot have intuitions of them, so one of the components required for cognition is missing. However, another way to understand Kant's claims that we cannot know things in themselves is in a stronger sense. In this stronger sense, Kant is claiming that we cannot have any knowledge at all about things in

---

[70] Jacobi noticed this very early on; see Jacobi, Friedrich Heinrich, *David Hume über den Glauben; oder Idealismus und Realismus* (Breslau: 1787).
[71] For an example of this strategy, see Rolf-Peter Horstmann, "Kants Paralogismen," *Kant-Studien* 84(4) (1993), pp. 408–425.
[72] See A 51/B 75.

themselves, including a priori knowledge or analytical knowledge. On this view, for example, we cannot know that things in themselves obey laws of logic. Thus there may be things in themselves that are squared-circles or things in themselves that possess some other contradictory properties. This is to say that we cannot even have basic analytical or a priori knowledge of things as they are in themselves. It appears that Kant argues only that we cannot have knowledge of things in themselves in the former sense rather than the latter.[73]

If we understand Kant's claims about knowledge of things in themselves in the latter sense, then we can understand how his claims regarding things in themselves may be legitimate even given our ignorance of things in themselves. One example here is that Kant argues that only appearances have spatial and temporal properties. Since things in themselves are not appearances, then they lack such properties. In order to make such an argument, however, Kant need not claim to cognize things in themselves using intuitions and concepts. Nevertheless, some facts about things in themselves follow from our a priori knowledge that space and time are pure forms of intuition and from our knowledge of appearances. This might be said to count as knowledge of things in themselves in some regard, but it is not knowledge in the sense of cognition through intuitions and concepts and so does not violate Kant's claims that we cannot cognize things as they are in themselves.[74] Although this is not an exhaustive argument for the reconciliation of Kant's seemingly contradictory claims, it does suggest that we should not be so quick to dismiss the passages in which Kant does make substantive claims about things in themselves and in particular passages in which he makes such claims regarding the ground of thought.

---

[73] The fact that Kant thinks laws of logic apply to things in themselves can be seen from the fact that he thinks the principle of complete determination applies to them, since complete determination also requires the principle of non-contradiction. On complete determination and things in themselves, see James Van Cleve, *Problems from Kant* (New York: Oxford University Press, 1999), pp. 224f.; see also A 571f./ B 599f. On complete determination and non-contradiction in Kant, see Nicholas F. Stang, "Kant on Complete Determination and Infinite Judgement," *British Journal for the History of Philosophy* 20(6) (2012), pp. 1117–1139.

[74] I do not address here interpretations that maintain that Kant holds that we cannot say anything meaningful about things in themselves. For an attempt to answer this kind of objection, see Colin Marshall, "Kant's Metaphysics of the Self," *Philosophers' Imprint* 10(8) (2010), pp. 5–6.

1.4 Conclusion

I have argued that we should attend to Baumgarten's a priori argument for the substantiality of the soul, which argues that the soul must be the ultimate substance in which the attributes of thought are grounded on the basis of the principle of sufficient reason and the claim that thought is an attribute. In doing so, we can then see that Kant understood the project of rational psychology and its arguments for the substantiality of the soul along the lines of Baumgarten's a priori argument. According to Kant, the rationalist argues that the ground of the attributes of thought is an absolute unconditioned subject. Kant argues, however, that the rationalist arrives at this conclusion on the basis of transcendental illusion. This illusion consists in thinking that the application of the principle of sufficient reason will lead to an ultimate unconditioned ground of the attributes of thought that is an appearance to which the spatial and temporal conception of substance applies. In contrast, Kant argues that the principle of sufficient reason will lead to a ground of the attributes of thought that is not an appearance but a thing in itself if it is anything at all. Since it is a thing in itself, the spatial and temporal conception of substance does not apply to it. And since the spatial and temporal conception of substance does not apply to this ultimate ground of thought, the rationalist cannot legitimately conclude that the soul is incorruptible, i.e. immortal. Kant does not, however, completely reject the idea that the ultimate ground of the attributes of thought is a substance. Kant appears to allow that the substance that grounds the attributes of thought is not a persisting spatiotemporal substance but a set of intrinsic properties, which are the powers of a substance to ground the attributes of thought. Although such properties may be logically abstracted from the substance to yield the notion of a bare substratum, this ground never exists as a bare substratum but rather only as the set of intrinsic properties or powers. It is unclear, however, what Kant thinks these powers might be. In the next chapter, we will consider Kant's discussion of the rationalist debate about the number and kind of powers the soul possesses and whether these powers could be reduced to a single *vis repraesentativa*, or power of representation.[75]

---

[75] On the common root of sensibility and understanding, see Dieter Henrich, "On the Unity of Subjectivity," in *The Unity of Reason* (Cambridge: Harvard University Press, 1994), pp. 17–54.

Chapter 2

Kant's Second Paralogism and the Powers of the Soul

2.1 Introduction

In the previous chapter, we considered Kant's confrontation in the First Paralogism with the Baumgartian a priori argument for the substantiality of the soul. It was concluded that Kant held that the application of the principle of sufficient reason to the attributes of thought implied that thought must be grounded in a thing in itself rather than an appearance. It was also concluded that Kant may have maintained that thought is grounded in the powers, or intrinsic properties, of a substance, but that this substance must not be construed as an enduring spatiotemporal substance, nor is the claim concerning its existence sufficient to demonstrate the immortality of the soul. Having seen that Kant may have maintained that thought is grounded in the powers of a substance, in this chapter we will consider Kant's discussions of the debates among his contemporaries about he number of powers the soul may possess. We will consider Kant's thoughts on this issue by looking at his lectures on metaphysics, the subjective deduction, and his discussion of the compositeness of the soul in the Second Paralogism of the first edition of the *Critique of Pure Reason* (1781), and we will illuminate his discussion of this issue by situating it in the broader context of German rationalism and responses to rationalist positions.

In the introduction to the *Critique of Pure Reason*, Kant suggests that the "two stems of human cognition […] namely sensibility and understanding" "may perhaps arise from a common but to us unknown root" (A 15/ B 29). But Kant's allusion to this fundamental power remains opaque, and it appears that it is not until much later in the *Critique*, in the Appendix to the Transcendental Dialectic, "On the regulative use of the ideas of pure reason," that Kant actually explicitly takes up the idea of a fundamental power from which the faculties of understanding and sensibility, or in other words spontaneity and receptivity, arise.[76] And here Kant suggests that it is merely a goal of reason to search for such a fundamental power or force but that we have no grounds for positing its existence. However,

[76] See A 649/ B 677. For other discussions of a fundamental power, see also: A 648/ B 676 – A 651/ B 679; A 682/ B 710 – A 684/ B 712; A 631ff./ B 659ff; A 771/ B 799.

the discussion of the number of mental powers also takes place in the Transcendental Deduction as well. Kant's talk of mental faculties and powers, particularly in the more psychologistic A edition of the *Critique*, is often rejected by interpreters as evidence of a legacy of faculty psychology in Kant's work that can be eliminated. Among those interpreters who take Kant's discussion of powers seriously from an historical point of view, there have been a few attempts to understand Kant's discussion of a fundamental power or force. Most notably, Dieter Henrich provides an account of the legacy of discussions of a fundamental power that Kant inherits form his German predecessors.[77] And more recently, Corey W. Dyck has also taken up Kant's discussion of a fundamental power arguing that faculty psychology is central to the subjective deduction and showing that Kant argues for the existence of multiple irreducible mental powers.[78] Both approaches suggest that the central place to uncover Kant's thoughts on a fundamental power that underlies our mental capacities (or powers) such as sensibility and understanding is in the discussion of psychology in the subjective deduction. This is without a doubt true, and Kant's disagreement with his predecessors is most explicitly expressed in this section. However, because interpreters have focused primarily on the subjective deduction and Kant's isolated statements on the fundamental power in the Appendix, some key elements of Kant's thought on a fundamental power are overlooked which promise to shed light on some metaphysical issues surrounding Kant's views on the kind and number of our mental powers.

A cursory look at the discussion of a fundamental power in the eighteenth-century German philosophical context suggests that for Wolff and Wolffian philosophers the issue of a fundamental power concerned not only the various powers and faculties humans exhibit in cognition but also fundamental metaphysical problems regarding the relationship between powers and the substances that support them. A central problem in this regard was whether the existence of a plurality of mental powers would entail that the soul itself is a composite. Thus the issue of a fundamental power has much deeper metaphysical roots than one might suspect when looking only at Kant's discussion of the issue in the subjective deduction or the Appendix. And Kant does not address the issue of whether a substance with multiple powers is a composite head on in the subjective deduction or the Appendix. But in the Second Paralogism Kant discusses the issue of the compositeness of the soul in such a way that

---

[77] See Dieter Henrich, "On the Unity of Subjectivity," in *The Unity of Reason* (Cambridge: Harvard University Press, 1994), pp. 19–40.
[78] See Corey W. Dyck, "The Subjective Deduction and the Search for a Fundamental Force," *Kant-Studien* 99(2) (2008), pp. 152–179.

provides a clue for understanding how he may have responded to some of the Wolffian arguments that held that the soul must have a single fundamental power. Such an approach to the Second Paralogism is admittedly at odds with how it is commonly read. On a commonly accepted reading of the Second Paralogism, Kant criticizes an unnamed rationalist for mistakenly concluding that the soul is simple on the basis of a flawed syllogism involving an ambiguous middle term. Kant argues against the rationalist that we have no epistemic justification for inferring that the soul is simple on the basis of formal features of the unity of apperception and therefore the conclusion that the soul is simple is unwarranted.[79] Although this argument is evident in Kant's discussion, the focus on the epistemological critique of the rationalist and the details of Kant's attempt to fit his critique of the rationalist neatly into the critique of a paralogistic syllogism overlook the fact that aspects of Kant's discussion of the rationalist's argument appear to be connected with larger a priori debates in metaphysics about the simplicity or compositeness of the soul and the issue of a fundamental power that continue the discussion raised in the subjective deduction.[80] A closer look at the debates among Kant's predecessors and Kant's responses to these debates in the subjective deduction, the Second Paralogism, and his lectures on metaphysics will provide a more thorough understanding of Kant's thoughts on a fundamental power.

In this chapter, I argue that Kant's Second Paralogism provides explicit and implicit resources for arguing against the Wolffian claim that the unity of thought and the nature of substances shows that the soul is simple and possesses a single fundamental power of representation. As such, the Second Paralogism also presents a continuation of Kant's

---

[79] See, for example: Michelle Gilmore Grier, "Illusion and Fallacy in Kant's First Paralogism," *Kant-Studien* 84(3) (1993), pp. 257–282; Ian Proops, "Kant's First Paralogism," *Philosophical Review* 119(4) (2010), pp. 449–495; Patricia Kitcher, "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), pp. 515–547; Graham H. Bird, "The Paralogisms and Kant's Account of Psychology," *Kant-Studien*, 91(2) (2000), pp. 129–145; Jill Vance Buroker, *Kant's Critique of Pure Reason: An Introduction* (Cambridge: Cambridge University Press, 2006), p. 221.

[80] Commentators disagree about the number of arguments Kant is discussing in the Second Paralogism. Norman Kemp Smith identifies three arguments for the soul's simplicity in *A Commentary to Kant's Critique of Pure Reason* (New York: Palgrave Macmillian, 2003), pp. 458–61. C. Thomas Powell considers five different arguments in *Kant's Theory of Self-Consciousness* (Oxford: Oxford University Press, 1990), chapter 3. However, the number of arguments is not very important. Arguments are individuated by sets of propositions, and there are simply too many propositions both expressed and unexpressed in the Second Paralogism to make counting arguments a worthwhile endeavour. The more important issue is to consider how the context of Kant's discussion may inform our understanding of the nature of the arguments he is making in the Second Paralogism.

argument for multiple mental powers in the subjective deduction and addresses some of the deeper ontological questions regarding powers and substances that troubled his predecessors. In my discussion, I proceed in the following way. In 2.2, I consider arguments provided by Christian Wolff for the thesis that our capacity for unified thought must be grounded in a single simple soul endowed with a single power, the implications of this view for proofs for the immortality of the soul, and the objections raised by Crusius and Lange to Wolff's arguments. In 2.3, I consider Kant's argument for multiple mental powers in the subjective deduction, indicate that the deduction leaves some questions regarding the ground of these powers unanswered, and show how resources for the answers to these questions can be found in the Second Paralogism. Section 2.4 concludes with a summary of the results of this chapter and raises questions that will be addressed in the subsequent chapters.

2.2 Wolff and Wolffians on the Powers of the Soul

It would seem that the most obvious place to look among Kant's historical predecessors when discussing the historical and dialectical background of Kant's discussion of the powers of the soul and the simplicity of the soul would be Baumgarten's *Metaphysica* (*Metaphysics*) (1739).[81] Kant lectured on the topic of metaphysics using Baumgarten's *Metaphysica* as a textbook throughout his career, even following his critical turn surrounding the publication of the *Critique of Pure Reason* in 1781, and a significant portion of these lectures was dedicated to the rationalist metaphysics of the soul and its understanding of powers and substances. However, Baumgarten's discussion of the soul does not present the detailed arguments for the simplicity of the soul that parallel those Kant critiques in the Second Paralogism, nor does Baumgarten discuss at length the debate about the number of powers with which the soul is endowed, which Kant discusses in his lectures on metaphysics, although Baumgarten does maintain that the soul is a single power of representation much as Wolff does.[82] A much more detailed discussion of the mental powers and their supporting substances required for thinking and the unity of thought can, however, be found in Wolff and responses to Wolff's arguments made by Christian August Crusius and Johann Joachim Lange.

In his discussions of the soul and its powers, Wolff relies on his views on the nature of thought established in *Vernünfftige Gedancken von Gott, der Welt und der Seele des*

[81] See Baumgarten, *Metaphysica* (Frankfurt: 1757), §745–747, §756, §757.
[82] See Baumgarten, *Metaphysica* (Frankfurt: 1757), §505–507

Menschen, auch allen Dingen überhaupt (*Rational Thoughts on God, the World and the Soul of Human Beings, Also All Things in General*), also known as the *Deutsche Metaphysik* (*German Metaphysics*) (1720).[83] According to Wolff, thinking, which he equates with consciousness, consists in the capacity to cognize the "difference between the soul and those things that are represented" (§730, §729), which also requires the capacity to distinguish between oneself and objects external to oneself and to differentiate among the various objects presented in conscious thought. He writes for example: "we find, accordingly, that we are conscious of things when we differentiate them from one another" (§729). We also become conscious of our thoughts of ourselves "when we notice the difference between ourselves and other things of which we are conscious" (§730). The capacity to distinguish between oneself and the objects of thought and between the various objects of thought also requires additional capacities. According to Wolff, "[i]f one wishes to distinguish things from one another, one must compare them" (§733). Comparison also requires the capacity to retain thoughts in memory: "When one compares the thoughts, one must not only retain what is thought but also know that one has already had these thoughts and so must have a capacity for memory" (§734).[84] So for Wolff, thought requires certain mental capacities, which include the capacity to retain thoughts, reflect on them [*überdencken*], compare them, and synthesize them into unified representations (§730, §733, §734, §735), all of which "is an activity [*Würckung*] of the soul" (§730). Because "a capacity [*Vermögen*] is only a possibility of doing something" (§117), these mental capacities are not sufficient for thought, so they must also be grounded in an actual power [*Kraft*] of the soul, which is the source of capacities and actual changes.

Wolff argues furthermore that the capacities associated with thought must be grounded in a single substance that possesses a single power.[85] Although it may appear that the soul has several powers that causally ground changes – for example, it has memories, imagination, desires, and other mental states that appear to ground various actions – Wolff argues that "a plurality of powers distinct from each other cannot be found in the soul,

---

[83] See Christian Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (*Rational Thoughts on God, the World and the Soul of Human Beings, Also All Things in General*) (*Deutsche Metaphysik*) (1720) (Halle: 1751). All translations of Baumgarten, Wolff, Knutzen, Lange, and Crusius are my own. I have sometimes consulted the translations in Eric Watkins (ed. and trans.), *Kant's Critique of Pure Reason: Background Source Materials* (Cambridge: Cambridge University Press, 2009).

[84] On Wolff, *Deutsche Metaphysik* §728, §730, §735–6, see Kant's Transcendental Deduction (A 84–A 130/ B116–B169).

[85] On Wolff's conception of the soul, see Richard J. Blackwell, "Christian Wolff's Doctrine of the Soul," *Journal of the History of Ideas* 22(3) (1961), pp. 339–354.

because otherwise every power would require a self-subsisting thing to which it would be ascribed" (§745). The fact that we appear to ourselves as having various capacities and powers is due to the fact that we are able to distinguish conceptually between these powers, but this does not entail that the these powers are also in fact distinct at a more basic ontological level (§745). A plurality of distinct powers cannot be found in the soul because, according to Wolff's ontology, any power must be grounded in a self-subsisting simple substance. And if the soul was or had several powers, then each power would need to be grounded in a self-subsisting simple substance, which would make the soul a composite. Wolff makes this argument very clearly in the *Psychologia rationalis* (1734), where he writes:

> *The power of the soul may only be a single one*. The soul is namely simple and therefore lacks parts. We may assume that the soul has multiple powers distinct from one another: if each of these consisted in a continuous striving for action, each of these would require a different subject in which it inheres. And so multiple actual beings each distinct from one another must be conceived, which, if it is assumed that they are the soul, are its parts, which has been demonstrated to be absurd. (§57).[86]

Wolff's point in the *Psychologia rationalis* is that because each power requires a separate and independent substance, if the soul had a plurality of powers, each of these would require a substance and therefore the soul would be a composite.

The idea that the soul could be a composite of substances each endowed with a different power is problematic for Wolff for a number of reasons, as can be seen from Wolff's arguments in the *Deutsche Metaphysik*. According to one argument, a power consists in a striving to do something, in an activity. If a soul consisted in several such strivings, then it would be pulled in different directions "as if a body, which is to be viewed in its motion as an indivisible thing (§667), should move in different directions at the same time" (§745), which is absurd.[87] A soul that was pulled in multiple directions could not exhibit the kind of unity characteristic of thought. Wolff also presents additional arguments intended to show that a composite would not be capable of thought (§738). When discussing a composite in the

---

[86] See Christian Wolff, *Psychologia rationalis* (Frankfurt: 1734); reprinted in Christian Wolff, *Gesammelte Werke* II/6, ed. Jean École (Hildesheim: Georg Olms, 1972).

[87] Wolff does not consider the possibility of two strivings that pull the soul in the same direction, but it may be argued on his behalf that either such strivings would pull entirely in the same direction, in which case they would be indiscernible and so actually be only one striving, or they would pull in different directions however slight, and so would indeed be different strivings. Wolff is also sceptical about how one would conceive of the interaction of multiple powers in a single substance; see *Psychologia rationalis* (Frankfurt: 1734), §57.

context of a composite of substances, Wolff equates compositeness with the kind of compositeness exhibited by matter. According to Wolff, if composite matter were to retain a thought across a period of time, then the parts that constitute the composite matter would have to be held in place or prevented from moving, since movement would cause an alteration in the composite and so also an alteration in the thought. Or these parts would need to be replaced with identical parts, which would allow the composite to retain the same arrangement of parts and so also to maintain the same thought.[88] Wolff also points out that if composite matter were capable of noticing a change in its thoughts, then it would have to be able to compare the differences and similarities between two states or arrangements of its body. This comparison, however, cannot be achieved "through the motion of parts." Wolff likely means by this that matter is incapable of reflection and synthesis because reflection takes another thought as an object and would thereby change the configuration of matter and therefore the thought. Importantly, Wolff's criticism of the ability of composite matter to produce thought applies to any composite of substances because he maintains that "no changes can occur in a composite thing other than in its size and shape, and the location of its parts, in its internal motion and in the place of the whole thing" (§72). So because Wolff maintains that each power requires one and only one substance, he thinks that any argument that accepts that the soul has more than one power will end up maintaining that the soul is a composite and so will also be susceptible to the kinds of arguments raised against materialist views that hold that matter is capable of thought.[89]

Wolff's argument against the idea that a composite could ground thought had a great deal of defenders and adherents eighteenth-century Germany such as Ludwig Philipp

---

[88] Wolff does not explicitly say why composite matter would not be capable of this, but similar points were often raised against materialist explanations of identity and persistence. Since matter is always changing, materialists have difficulty explaining, for example, the persistence of persons or objects across time. Likewise, since matter is always changing, it is not immediately evident that the materialist will be in a position to explain how a thought, which is nothing but a certain arrangement of matter, can be retained for anything but the briefest period of time.

[89] Locke's claim that God could superadd the power of thought to matter was a matter of great controversy. See John Locke, Locke, *An Essay Concerning Human Understanding*, ed. P. Nidditch (Oxford: Clarendon Press, 1975), IV.iii.6. On Locke's thesis and the surrounding debates, see John W. Yolton, *Thinking Matter: Materialism in Eighteenth–Century Britain* (Oxford: Blackwell, 1984) and John W. Yolton, *Locke and French Materialism* (Oxford: Clarendon Press, 1991).

Thümmig and Georg Bernhard Bilfinger.[90] In his *Philosophische Abhandlung von der immateriellen Natur der Seele* (*Philosophical Treatise on the Immaterial Nature of the Soul*) (1744), however, Kant's teacher Martin Knutzen objects that Wolff's argument against the ability of a composite to produce thought assumes that the powers of each component substance of the composite are a *vis motrix* or power of motion, which leads Wolff to claim that the proponent of a composite must hold that thought could come about through a new arrangement of its parts or location through motion (§738). Since the proponent of the claim that composite matter is capable of thought might reject the reduction of the powers of the substances in the composite to a *vis motrix*, and instead argue that a composite could be endowed with representational powers, Knutzen argues that a revised argument is needed. In his presentation of an alternative argument in §6 of his *Philosophische Abhandlung*, Knutzen essentially adopts the Wolffian view of thought, suggesting that "whatever thinks must be conscious of itself and other things" and that consciousness requires the capacity to distinguish oneself from other things (§2).[91] According to Knutzen, this capacity to distinguish oneself from other objects requires that all representations be grounded in a single thinking thing that is capable of comparing and synthesizing representations (§3). This single thinking thing cannot, however, be composite matter. This is because the representations that are distinguished from one another must be "pictured," compared, contrasted, and synthesized in a single subject, and this activity must come about through a single "efficacious power." As a composite, however, matter "consists of actual parts, or of parts that are posited external to each other," is "nothing other than the sum or connection of parts," and consists of several efficacious powers or causal grounds (§4). The unity of thought that arises from the activity of comparing and synthesizing representations cannot come from the collective action of several discrete parts, each endowed with a distinct power. And since composite matter is incapable of this activity, it is incapable of thought. Instead, Knutzen proposes that the unified and unifying activity associated with thought must be grounded in a single substance, "which is completely devoid of all parts" (§4), and is endowed with a single power of thought.

---

[90] See Ludwig Philipp Thümmig's *Institutiones philosophiae Wolfianae* I (Frankfurt: 1725), §174, and Georg Bernhard Bilfinger's *Dilucidationes philosophicae* (Tübingen: 1746), § CCLXXI. For a discussion of 18th century objections to Wolff, see Falk Wunderlich, *Kant und die Bewußtseinstheorien des 18. Jahrhunderts* (Berlin: Walter de Gruyter, 2005), pp. 40–46.

[91] See also Martin Knutzen, *Philosophische Abhandlung von der immateriellen Natur der Seele* (*Philosophical Treatise on the Immaterial Nature of the Soul*) (Königsberg: 1744), §7.

It is important to note, however, that regardless of the relative merits of each argument against the ability of a composite to think, both Wolff and Knutzen maintain that any argument that would maintain that the soul is endowed with a plurality of powers would entail that the soul is a composite of spatial substances. And since the soul is a composite it would be subject to the kinds of arguments against the ability of composite matter to think noted above. As a composite, a soul endowed with multiple powers would simply not be capable of producing a unity of thought. Moreover, for Wolff and Wolffians such as Knutzen, there are also additional reasons for arguing that the soul cannot be a composite. The most important of these is that if the soul is composite it would be subject to a dissolution of its parts and so could not be immortal. In order to demonstrate the possibility of an afterlife, they argue instead that the soul is a simple substance endowed with a single power of representation capable of producing a unity of thought and incapable of perishing through a dissolution of its parts. Problematically, however, in their arguments for both the incapacity of composite matter to think and the incorruptibility of the simple soul, Wolff and Wolffians such as Knutzen share the foundational premise that a distinct power must be grounded in a single distinct spatial substance. For it is only of one accepts this premise that one would think that a soul endowed with multiple powers must be a composite like matter and so would be incapable of thought and subject to a dissolution of its parts.

Wolff's idea that thought and the capacities for thought must be grounded in a single power of a single substance, and the implications of this view for the simplicity of the soul and immortality, met with criticism from some of his contemporaries including Crusius and Lange. In his *Entwurf der nothwendigen Vernunft-Wahrheiten* (*Sketch of the Necessary Truths of Reason*) (1745), the Pietist philosopher Crusius argues in favor of the idea that the soul can have a plurality of powers without being a composite. He writes:

> The understanding of a rational but finite spirit is not a single fundamental power [*Grundkraft*]; rather, one must think of it as a totality [Inbegriff] of certain fundamental powers [*Grundkräfte*] and certain powers [*Kräfte*] and capacities [*Vermögen*] derived from them, which all have this in common, that they consist in a mode [*Art*] of thought and together contribute to the promotion of the knowledge of truth accompanied by consciousness. (§434, p. 907)[92]

For Crusius, these powers are taken together to contribute to the overall capacity of a finite being for consciousness and the knowledge of the truth. Although we cannot know with

---

[92] See Christian August Crusius, *Entwurf der nothwendigen Vernunft-Wahrheiten* (1745) (Leipzig: 1766). Page numbers from Crusius refer to this edition.

certainty how many such powers there are, we would be acting contrary to the evidence if we were to posit that the reason of a finite spirit is grounded in a single power as Wolff does.

Crusius also presents a couple of arguments against the Wolffian thesis that the soul can have only one power. One argument goes as follows.

> First one must consider that every idea is an action, and also that in an idea something different is always represented from that in another idea, therefore, I conclude, the proximate actions of a fundamental power [*Grundkraft*] would not consistently be similar, which must be the case if all ideas were activities of a single fundamental power. (§434, p. 909)

According to Crusius, our ideas and mental states exhibit a great deal of qualitative difference. If, however, one thinks as Wolff does that each of these mental states is grounded in a single fundamental power, then a problem arises. Since each idea is qualitatively different, it shows that the fundamental power is inconsistent in the effects it produces. But this seems contrary to what one would expect of a fundamental power. One would expect that a fundamental power would produce consistent effects. The fact that one needs to account for the variety and qualitative difference of ideas might then lead one to think that they must be grounded in a plurality of powers. Although Crusius admits that not every idea would need a power, since some ideas are constructed out of other ideas and thus could be a composite of powers, he nevertheless suggests there is good reason to think that the soul is endowed with more than one power. A second argument also relies on the variety of our mental capacities to argue that they must be grounded in more than a single power. He writes:

> Through consciousness [*Bewußtsein*] we ourselves have a representation of our own thoughts. The sun itself and the representation of the sun are the same thing just as little as the idea of the sun, i.e. the action through which it is thought, is the same thing as the representation through which the previous action itself is thought. For just as the sun is the object of the idea of the sun, so too is the idea of the sun in consciousness also the object of that idea through which it itself is represented and thought [So ist die Idee von der Sonne beym Bewußtsein wiederum das Object derjenigen Idee, wodurch sie selbst vorgestellet und gedacht wird]. One must therefore admit that consciousness requires a special fundamental power [*Grundkraft*] through which it is possible. (§434, p. 910)

According to Crusius, we should reject the Wolffian idea that consciousness arises through comparison and instead recognize that consciousness involves a second-order reflection on one's representations. If we adopt this understanding of consciousness, however, we see that the capacity to represent something and the capacity to be conscious of this representation

must require two different fundamental powers, which suggests that the soul must be endowed with at least two powers.[93]

Crusius also argues against the central Wolffian thesis that a soul endowed with multiple powers would be a composite. He writes:

> Incidentally, I would not be at fault if someone wished to conclude from the claimed different powers [*Kräften*] and actions of the soul that the soul is composite. For one need not imagine an idea as a particular substance nor as a particular motion, which must occupy their particular little parts or spaces in the substance. These would all be materialist concepts, which have already been refuted (§435). If one discards these, and does not seek to think anything material in an idea, then a composite of substances does not follow from the manifold of spiritual powers and their actions; rather, only a manifold activity and a perfection of the subject and its essence that exceeds that of matter follows. (§434, p. 913)

According to Crusius, one need not conclude that the soul is a composite from the fact that it possesses a number of powers. This is because one need not think as Wolff does that each distinct power requires a distinct substance, or that these powers and the individual substances that ground them occupy some quadrant of space. To do so would be to think of powers along the lines of matter. Crusius suggests that what follows from the assumption of a plurality of powers in the soul is not a composite of substances but a plurality of action or activity. One substance can produce a variety of actions through its plurality of powers. Importantly, given this view of the powers of the soul, Crusius may accept both that mental capacities and thought are grounded in a plurality of powers and that the soul is a simple, non-composite substance. Since the possession of multiple powers does not entail that the soul is composite, it may be regarded as simple. And since the soul is simple, it is immortal in the sense that it cannot be corrupted through a dissolution of its parts. Similarly, Lange also objects to Wolff that one need not think that each power requires a distinct substance, and he suggests that the composition of powers of the soul therefore need not be thought of mereologically in terms of the composition of spatial parts, which also allows Lange to maintain that the soul has multiple powers and is nevertheless simple and incorruptible.[94]

---

[93] In this argument, Crusius is also sceptical of the Wolffian claim that consciousness arises through the distinction of oneself form other objects. See Crusius, *Entwurf der nothwendigen Vernunft-Wahrheiten* (1745) (Leipzig: 1766), p. 911. He believes that differentiation requires consciousness and not that it results in consciousness as Wolff argues. So Crusius argues against Wolff's grounding of the capacities of thought in single power by attacking the notion of consciousness that underlies Wolff's conception.

[94] See Joachim Lange, *Modesta disquisitio novi philosophiae systematis de deo, mundo et homine* (Halle: 1723), p. 72, p. 68; reprinted in Christian Wolff, *Gesammelte Werke* III/23, J. Lange, *Kontroversschriften gegen die Wolffische Metaphysik* (Hildesheim: Georg Olms,

These discussions regarding the powers of the soul were well known to Kant, so it is not surprising to find Kant engaging directly with both the Wolffian arguments and Crusius' criticism of these arguments in his lectures on metaphysics and the *Critique of Pure Reason*.

2.3 Kant on Mental Powers and the Soul

2.3.1 Powers and the Transcendental Deduction

In the previous section, we saw that the discussion of the number of powers that could be attributed to the soul was intimately connected with debates about whether a single power must be grounded in a single substance, whether multiple powers or capacities could be grounded in a single substance, whether the existence of multiple powers would entail a composite soul, and whether a composite soul might be accepted. Kant's multiple statements throughout the *Critique of Pure Reason* and the lectures on metaphysics indicate that he was well aware of the connections between questions about the number of mental powers and the nature of the soul that grounds these powers.[95] Kant also takes up the issue of the number of mental powers explicitly in the subjective deduction. He suggests in the preface to the A edition of the *Critique of Pure Reason* that the purpose of the subjective deduction is to consider "How is the faculty of thinking itself possible?" (A xvii). Kant glosses this as "something like the search for the cause of a given effect." This as we saw was exactly how Wolff raised the question of a fundamental power, by considering the faculties and powers that are required in order for thinking to be possible. Similarly, Kant's aim in the subjective deduction is to provide an account of the various powers that give rise to or cause our capacity for thinking. In contrast with Wolff, however, who maintains that all powers of thought that we exhibit and the various capacities needed for synthesis are grounded in a single representative power (*vis repraesentativa*), Kant argues that our mental capacities, particularly sensibility and understanding, cannot be reduced to a common cause or power

1986). See also Joachim Lange, *Anmerckung über des Herrn […] Christian Wolffens Metaphysicam* (Kassel: 1724); reprinted in Christian Wolff, *Gesammelte Werke* I/17, *Kleine Kotroversschriften mit Joachim Lange und Johann Franz Budde*, ed. Jean École (Hildesheim: Georg Olms, 1980).

[95] See for example Metaphysik L_1, AA 28:261–62.

but that the mental powers he identifies in the subjective deduction are irreducible and jointly necessary for cognition.[96]

Regarding the manner in which he argues for the existence of these irreducible and jointly necessary mental capacities, in a paragraph omitted from the second edition of the *Critique*, Kant writes:

> There are, however, three original sources (capacities or faculties of the soul), which contain the conditions of the possibility of all experience, and cannot themselves be derived from any other faculty of the mind, namely sense, imagination, and apperception. On these are grounded 1) the synopsis of the manifold *a priori* through sense; 2) the synthesis of this manifold through the imagination; finally 3) the unity of this synthesis through original apperception. In addition to their empirical use, all of these faculties have a transcendental one, which is concerned solely with form, and which his possible *a priori*. (A 94)

Unlike Crusius, who rejects Wolff's account of consciousness, Kant does not appear in the *Critique of Pure Reason* to disagree with Wolff's idea that coherent and unified thought requires the capacity to compare and synthesize thoughts.[97] Indeed, Kant sets out three kind of synthesis – of the manifold through sense, through the imagination, and the unity of this synthesis through apperception – that contribute to the possibility of experience, by which Kant means the unity exhibited in our thinking. Although Kant and Wolff differ in the details about how to classify these kinds of synthesis, they are in broad agreement for example that

---

[96] Unlike Wolff, who makes a strict distinction between a faculty (*Vermögen*) and a power (*Kraft*), Kant does not adhere closely to this distinction. Thus he alternately refers to a capacity to judge (*Vermögen zu urteilen*) (A 69/B 94), or equivalently a capacity to think (*Vermögen zu denken*) (A 81/B 106), and the power of judgment (*Urteilskraft*) (A 136/B 175). However, Kant was well aware of Wolff's distinction, and it appears likely that he would not have objected to the idea that a power is needed in order for a faculty to be exercised. In Metaphysik Volckmann Kant writes for example: "Capacity [*Vermögen*] and power [*Kraft*] must be distinguished. In capacity we represent to ourselves the possibility of an action, it does not contain the sufficient reason of the action, which is power [Kraft], but only its possibility" (AA 28:434). See also Metaphysik Mrongovius AA 29:822ff. and Metaphysik L_2 AA 28:565 for Kant's discussion of Wolff's distinction between faculty and power. Beatrice Longuenesse also suggests that Kant is sometimes but not always strict in making this distinction in the *Critique of Pure Reason*; see Longuenesse, *Kant and the Capacity to Judge* (Princeton: Princeton University Press, 1998), pp. 7–8.

[97] It has been suggested that Moses Mendelssohn's view of thought may have been influential for Kant's discussion of the unity of thought in the *Critique of Pure Reason*, but it is much more likely that the elements of Mendelssohn's view enlisted to support this claim – the synthesis of representations into a unity, in particular – have their root in the Wolffian conception of thought which was so influential during the period in which Kant was writing. See Brigitte Sassen, "Kant and Mendelssohn on the Implications of the 'I think'," in *The Achilles of Rationalist Psychology*, ed. T.M. Lennon and R.J. Stainton (Dordrecht: Springer 2008), p. 228.

representations must be combined and that this requires some kind of retention of representations and their ascription to a single subject. However, rather than suggesting that such forms of synthesis can be reduced to a single power of representation, Kant argues that these forms of synthesis have "three original sources (capacities or faculties of the soul) [*Fähigkeiten oder Vermögen der Seele*], which contain the conditions of the possibility of all experience, and cannot themselves be derived from any other faculty of the mind, namely sense, imagination, and apperception" (A 94).

It is not important for our purposes here to uncover the details of how Kant argues that these three transcendental faculties are irreducible to a single fundamental power and are jointly necessary for thinking.[98] Suffice it to say that Kant identifies certain empirical capacities that are used in cognition and proposes that each has a necessary transcendental ground. It is only important to note that regardless of whether the argument is convincing or not, there is a deep component of the historical discussion of mental powers that is in part left out of Kant's discussion in the subjective deduction, namely the issue of the substance in which the powers reside. As can be seen from the passage above in which Kant announces the aims of the subjective deduction, he suggests that the "three original sources," "which contain the conditions of the possibility of all experience," are "capacities or faculties of the soul" (A 94). Although Kant's wording here regarding a soul might appear merely to be a *façon de parler*, it is not. Rather, it suggests that although Kant explicitly disagreed in the subjective deduction with the reduction of mental powers to a single fundamental power, he nevertheless recognized that the discussion of mental powers was intimately tied to discussions about the powers of the soul. And this is not surprising given Kant's familiarity with the debates about the powers of the soul. Kant's allusion to the soul also reflects his well-known ambivalence throughout the A edition regarding the substantiality of the soul. Nor is it, however, surprising that Kant does not engage explicitly with the debate about whether a soul could possess multiple powers in the subjective deduction since the focus of the Deduction and the entire Analytic of Concepts in the Transcendental Analytic is not primarily with the metaphysical views of his predecessors but attempts as much as possible to bracket such discussions in order to develop an analysis of cognition.

---

[98] Corey W. Dyck provides such an argument in Dyck, "The Subjective Deduction and the Search for a Fundamental Force," *Kant-Studien* 99(2) (2008), pp. 152–179. Although I agree with the broad outlines Dyck's interpretation of Kant's argument for multiple powers, I maintain that Kant's argument leaves some questions unanswered that are later discussed in the Second Paralogism.

However, given the immense depth of the questions raised by Kant's predecessors, the discussion of a fundamental power in the subjective deduction leaves a number of questions unanswered. First, the subjective deduction leaves open the obvious objections raised by Kant's predecessors to a plurality of mental powers. It is unlikely that Kant would have held that mental powers are not grounded in anything more substantial, and since he appears at least to suggest that these capacities are indeed capacities of the soul, an adequate demonstration of the possibility of a plurality of mental powers would need to answer the objections raised by Wolff and others against multiple powers. Second, the subjective deduction does not explain why Kant in the preface suggests that the faculties of sensibility and understanding "may perhaps arise from a common but to us unknown root" (A 15/ B 29). If the subjective deduction has shown that we have certain mental capacities that are irreducible to a fundamental power and are jointly necessary for our cognition, then it is unclear why Kant would nevertheless maintain that our faculties might possibly be grounded in some "unknown" power. If the argument of the subjective deduction along with its support in the objective deduction is decisive, as Kant apparently thinks it is, then it demonstrates that if we have the kind of mental life that we do, there can be no fundamental power at the basis of the faculties that make this mental life possible.[99] So it is deeply mysterious why Kant believes there is anything mysterious about the existence or non-existence of a fundamental power.

It appears that neither of these questions has an answer within the framework of the Transcendental Deduction. The most likely explanation for this is that the questions that are raised regarding the fundamental powers are questions that Kant reserves for treatment within the context of the discussion of rational psychology and transcendental idealism in the Transcendental Dialectic. Kant first discusses the question of whether the soul can be a composite, which is something that Wolff would argue is entailed by Kant's claim that the soul has a plurality of fundamental powers, in the Second Paralogism. The fact that a number of questions that need to be answered in order for Kant's account of multiple powers of the mind in the subjective deduction to be satisfying suggests that there is some continuity between the account of mental powers in the subjective deduction and the Second Paralogism

---

[99] Although Kant concedes that the subjective deduction is hypothetical and involves providing a description of our mental powers, he argues that the results of the deduction are not mere opinion but are further supported by the results of the objective side of the Transcendental Deduction; see A xvi–xvii.

such that they might be regarded as two aspects of an argument for multiple powers. It also appears that some answer to the question regarding why it is so mysterious whether there is a fundamental power can be provided by looking at Kant's account of the soul in the Paralogisms. In the Paralogisms, Kant comes to reject some of the rationalist's claims about the soul and our ability to cognize the soul as it is in itself. It appears that Kant may have thought that although it might be argued that certain mental powers are required for our thought, the question of whether these powers are grounded in a fundamental power is as difficult as any question regarding the nature of things in themselves.

2.3.2 Powers and the Simplicity of the Soul

Kant's emphasis on the mental powers required for the unity of thought is also carried over into the discussion of thinking and the unity of the subject of thought in the Second Paralogism. In the Second Paralogism, Kant again agrees with the Wolffian idea that thought requires the synthesis of representations into a unity. He concedes, for example, that "we demand absolute unity for the subject of thought" (A 354). What Kant means by this is that representations can count as the representations of one thing only if they are actively synthesized in judgments and attributed in each case to the same I, a point that he has demonstrated in the Transcendental Deduction.[100] However, although this idea that the unity of thought requires a synthesis of the components of thought and their ascription to a single I may on the surface appear very close to the Wolffian view, Kant ultimately argues that the Wolffian thesis that the unity of thought must be grounded in a simple substance fails.

In his discussion in the Second Paralogism regarding why the Wolffian thesis fails, Kant first considers the Wolffian idea that a composite of substances cannot produce thought as an effect. In his explication of the argument for this Wolffian thesis, Kant first considers

---

[100] Commentators disagree about the relevance of the Transcendental Deduction for the Paralogisms. Patricia Kitcher argues that it plays an important role in "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), pp. 515–547. Grier disputes this claim to some degree; see Michelle Grier, *Kant's Doctrine of Transcendental Illusion* (New York: Cambridge University Press, 2001), p. 167. I suggest that it is important to recognize that Kant accepts certain Wolffian conceptions of the mind in the Transcendental Deduction and argues against some of the metaphysical implications of the Wolffian view of the mind in the Paralogisms.

what a composite is and then distinguishes between a composite that produces an external effect as an accident and a composite that produces an internal effect as an accident.[101]

> Every composite substance is an aggregate of many, and the action of a composite, or that which inheres in it as such a composite, is an aggregate of many actions or accidents, which is distributed among the multitude of substances. Now of course an effect that arises from the concurrence of many acting substances is possible if this effect is merely external (e.g., the movement of a body is the united movement of all its parts). Yet with thoughts, as accidents belonging inwardly to a thinking being, it is otherwise. For suppose that the composite were thinking; then every part of it would be a part of the thought, but the parts would first contain the whole thought only when taken together. Now this would be contradictory. For because the representations that are divided among different beings (e.g., the individual words of a verse) never constitute a whole thought (a verse), the thought can never inhere in the composite as such. Thus it is possible only in one substance, which is not an aggregate of many and hence it is absolutely simple. (A 351–2)

According to Kant, the rationalist recognizes that a composite of substances may causally ground a unified effect when this effect is merely external. This is the case with any physical body, where its overall movement is grounded in the movement of each of its parts. For example, the parts of a human body, its legs, arms, and so on are each a substance that produces an action through its power: the muscles of the legs become tense and release, the arms move forward and backward. And these individual actions each combine to produce the effect or activity of walking. Using Kant's terminology of accidents, walking can also be an attribute of a composite, as in the statement 'the man is walking', where the accident 'walking' is attributed to the composite substance 'man'. The rationalist denies, however, that this is also the case with a thinking being our soul, where the effect, thinking, is internal rather than external. Kant illustrates this point with the so-called "Achilles argument" using the analogy of a verse.[102] Imagine that the individual words of a verse were divided among several individuals. If this were the case, the individual words of the verse would not constitute a whole verse since there would be no single individual who would be aware of the whole verse. As we have seen, for example, Knutzen argues that the various parts of a thought could not be synthesized into unified thought unless this synthesis is grounded in the efficacious power of a single substance. Similarly, Kant's rationalist argues that "the

---

[101] On the distinction between external and internal effects, see Falk Wunderlich, "Kant's Second Paralogism in Context," in *Between Leibniz, Newton and Kant*, ed. W. Lefevre (Netherlands: Springer, 2001), p. 180.

[102] Earlier discussions of the verse argument can be found in Metaphysik Herder (AA 28:44). See also *Dreams of a Spirit–Seer*, AA 2:322, AA 2:328n. Knutzen actually provides such a verse argument in *Philosophische Abhandlung* (Königsberg: 1744), §7–8.

representations that are divided" among a composite cannot "constitute a whole thought" because the unity required for thought cannot be causally grounded in a composite. And since the thoughts cannot have their causal ground in a composite, they must have a causal ground in an "absolutely simple" substance.

Kant is clear that there are no good epistemic reasons for holding that the *nervus probandi* of the argument is a warranted synthetic a priori or a posteriori truth since the ground of the unity of thought cannot be an object of cognition. According to Kant, we cannot legitimately infer that the soul is simple from the fact that there is a unity of thought since the proper condition for the application of the schematized concept of unity is lacking. This also means that the conditions are lacking under which it would be possible to have knowledge of the simplicity of the thinking self. This suggests that our epistemic limitations prevent us from knowing anything about the simplicity or compositeness of the ground of the unity of thought. Kant's epistemological response is, however, not particularly interesting or innovative since Locke and other empiricists had already argued that the nature of the substance underlying the unity of thought cannot be cognized (i.e. that we can know neither synthetic a priori nor a posteriori truths about it).[103] Moreover, although it is true that Kant denies that we can cognize whether the ground of thought is simple or not, such an epistemological argument does not respond directly to the rationalist. As we have seen neither the Wolffian rationalist nor Kant's rationalist claim to have empirical cognition of the simplicity of the soul but claim only that an a priori argument shows that it is the case that a composite cannot ground the unity of thought.

Beyond Kant's rejection of the Wolffian thesis as an a posteriori or synthetic a priori truth, however, he also attacks the rationalist conclusion on its own terms by arguing that it is neither an analytic a priori nor a necessary truth that the unity of thought cannot be produced by a composite of substances. He writes:

> [T]he unity of a thought consisting of many representations is collective, and, as far as mere concepts are concerned, it can be related to the collective unity of the substances cooperating in it (as the movement of a body is the composite movement of all its parts) just as easily as to the absolute unity of the subject. Thus there can be no insight into the necessity of presupposing a simple substance for a composite of thought according to the rule of identity. (A 353)

---

[103] See John Locke, *An Essay Concerning Human Understanding*, ed. P. Nidditch (Oxford: Clarendon Press, 1975), IV.iii.6.

Similar to the example of the walking man, Kant maintains that it is logically possible that the unity of thought as an internal attribute is grounded in a composite of distinct substances acting together to produce the unity of thought. One might think for example of the statement 'the man is thinking'. Since there is nothing incoherent or contradictory about this idea, it cannot be an analytic a priori or necessary truth that thought cannot be grounded in a composite. This relies on the simple modal truth that it is possible that *X* if and only if it is not necessary that not *X*. Likewise, if it is possible that thought can be grounded in a composite of substances, then it is not necessary that thought cannot be grounded in a composite of substances. However, although Kant points out this possibility, he is reticent about how such a possibility could be explained. Recall that because Wolff thought that a composite must produce its effects through an arrangement of its parts, thought could not be produced by a composite. And Knutzen argued that multiple efficacious powers could not work together to produce thought. But it is quite open to Kant to argue that it is logically possible that the unity of thought is an emergent property that has its ground in a composite of substances. Kant was likely well aware of the postulation of emergent properties in chemistry by seventeenth and eighteenth century scientists who held, for example, that the properties of water could not be derived solely from the properties of hydrogen and oxygen, so it would not be anachronistic to think that Kant may have had emergent properties in mind when he dismisses the rationalist's argument.[104]

---

[104] Kant's claim, however, that the unity of thought is an emergent property would have faced some additional counterarguments. Crusius provides one such argument in the context of a refutation of the idea that thought could be the effect of the power of motion or a *vis motrix*. According to Crusius, to say that thought could be the effect of a motion would be to ascribe an effect to a cause that is incompatible with the nature of the cause and so logically contradictory. Accordingly, movement can only produce a movement or a disposition to movement and not concepts and thoughts since thought has more perfection than movement. Similarly, one might argue that a ground that does not itself possess thought cannot produce thought because to do so would be to posit more perfection in the effect than in the cause. See Crusius, *Entwurf der nothwendigen Vernunft-Wahrheiten* (1745) (Leipzig: 1766), §429 and §430. For his argument that matter is incapable of thought, see §473. Other philosophers such as Moses Mendelssohn in his *Phädon* (1767) present similar arguments that might be enlisted to undermine the notion of thought as an emergent property. Mendelssohn also argues that properties such as harmony, beauty, and symmetry cannot be emergent since thought is required to produce such properties. Since the unity of representations arises from thought, the capacity to unify cannot itself be an emergent property. See Moses Mendelssohn, *Phädon oder Über die Unsterblichkeit der Seele* (Berlin: 1767); translated into English as *Phaedon, or the Death of Socrates*, trans. C. Cullen (London: 1789), pp. 121–138.

Kant's argument against the claim that the unity of thought cannot be grounded in a composite of substances is also interesting for the question of the number of powers the soul may possess. In Metaphysik L_1, Kant expresses one line of the Wolffian argument for a fundamental power as follows:

> Now in order to answer and to treat the question, whether all forces of the soul can be derived from one basic force, or whether several of them are to be assumed, we must of course say: because the soul is indeed a unity, which will be demonstrated later, and which the I already proves, then it is obvious that there is only one basic force, out of which all alterations and determinations arise. (AA 28:261–62)

The argument Kant attributes to the Wolffian maintains that a multiplicity of powers would violate the unity of the soul. It is thought that since our thinking exhibits unity there can be only one basic power in the soul. As we have seen, one essential component of the Wolffian position is that each power must be grounded in a distinct and independent substance. So, any view that posits a plurality of mental powers contributing to the unity of thought entails that the soul is a composite. Kant's argument, however, demonstrates that the fact that there is a unity of thought expressed in the "I" or "I think" does not prove anything about whether the ground of this thought is simple or composite. So even if one accepts, as the Wolffian does, that each power requires a distinct substance, this does not rule out the possibility that the unity of thought could be grounded in a composite of distinct substances each endowed with a distinct power.

However, although Kant's argument already shows that the rationalist cannot legitimately argue from the unity of thought, exhibited by the "I," to the simplicity of its ground and then to the existence of a single fundamental power, Kant also rejects the foundational premise upon which the Wolffian argument is built, namely the claim that each power must be grounded in a distinct and independent substance. Much like Crusius, Kant is skeptical of the a priori arguments for the claim that a substance may possess only one power and any substance that possesses more than one power is a composite. Throughout his lectures on metaphysics, Kant expresses skepticism of the Wolffian view. In Metaphysik Herder (1762–1764), for example, he is reported as saying:

> Each substance has powers [*Kräfte*]: it can have many fundamental powers [*Grundkräfte*] without being composite because the plurality of the accidents does not make the substance itself composite. The soul has many powers. (AA 28:29)

And he quite explicitly criticizes the Wolffians when he writes:

> The Wolffians falsely assumed that the soul qua simple has merely one power [*Kraft*] of representation. This arises because of an incorrect definition of power: because it is

> merely a respectus, the soul can have many respectus. As various as the accidents are that cannot be reduced to another. (AA 28:145)

As we have seen, the Wolffians think of powers as residing in a substance and occupying the same region of space as the substance. This means that anything that has multiple powers has spatially distinct parts, i.e. substances endowed with powers, of which it is composed. Since it is composed of spatially distinct parts, the Wolffian holds that such a composite is no different than composite matter. In the passages from the lectures on metaphysics, Kant argues against this spatial and mereological conception of powers and the substances in which they reside. He argues instead that a power is merely a property of a substance. Just as an object can have many properties, a substance can have many powers. And because these powers are merely properties, the fact that a substance has many properties does not entail that the substance is a material composite with distinct spatial parts in which each power resides. Kant does not appear to abandon his criticism of the Wolffian thesis that a soul with multiple powers would be a composite in the *Critique of Pure Reason* either. Indeed, Kant's discussion in the Second Paralogism suggests an additional argument that may be made against the Wolffian thesis by enlisting the distinction between appearances and things in themselves. Although Kant himself does not explicitly make such an argument against the Wolffian, it is nevertheless worth considering because it sheds light on how Kant may have developed his claims in the subjective deduction that thought requires multiple irreducible and jointly necessary powers.

In order to understand the Kantian argument that may be raised against the Wolffian, it will help to look briefly at Kant's criticism of the rationalist claim that simplicity entails imperishability. At A 356–361, Kant points out that the sole reason the rationalist wishes to establish that the soul is simple is in order to distinguish it from matter, which is composite. The motivation here is that if the soul can be shown to be non-composite, then it entails that it cannot perish through a dissolution of its parts. As Kant writes: "[T]he assertion of the simple nature of the soul is of unique value only insofar as through it I distinguish this subject from all matter, and consequently except it from the perishability to which matter is always subjected" (A 356). According to Kant, properties associated with matter such as compositeness, extension, and motion are properties only of appearances of outer sense and not of things in themselves. Whatever it is that grounds thought as a thing in itself, however, would not have such properties. As he writes: "But this Something is not extended, not impenetrable, not composite, because these predicates pertain only to sensibility and its

intuition, insofar as we are affected by such objects (otherwise unknown to us)" (A 368). The ground of thought may according to Kant be simple although it appears as a composite. He writes:

> [H]ence I can well assume about this substratum that in itself it is simple, even though in the way it affects our outer sense it produces in us the intuition of something extended and hence composite; and thus I can also assume that in the substance in itself, to which extension pertains in respect of our outer sense, thoughts may also be present, which may be represented with consciousness through their own inner sense. In such a way the very same thing that is called body in one relation would at the same time be a thinking being in another […]. (A 360)

According to Kant, the noumenal substratum of thought may be simple although it affects us in such a way that it produces an appearance that is extended and composite. Moreover, it appears that given the passage at A 368 and A 360, Kant maintains that the property of compositeness may not apply to this substratum because this property applies only to sensible things that we intuit, by which Kant means that compositeness applies only to appearances. In this regard, Kant is thinking of compositeness only as a property of the appearance of extended matter. Kant argues on the basis of the restriction of the property of compositeness to extended matter in appearance, that it may be the case that the noumenal ground of the appearance of matter may be simple although the effects of this noumenal ground are extended and composite. This is to say that just as a composite may produce a unity as an emergent property, so too may something simple produce a composite appearance that emerges when we are affected by a noumenal substratum. This view of the property of compositeness allows Kant to grant the rationalist that the ground of the unity of thought could be simple and so capable of thought and nevertheless coherently maintain that the simple soul may appear as composite matter.[105] Kant in turn also maintains that the soul could perish through a dissolution of its parts even if it is simple. This is because it may be a composite as appearance, and so subject to a dissolution of its parts, and simple as it is in itself.

Kant's argument against the idea that simplicity entails incorruptibility, which he makes on the basis of the observation that the kind of compositeness associated with matter applies only to appearances, is without a doubt terrible. For one thing, it does not really answer the rationalist to claim that the soul may perish through dissolution in one regard but in another regard may not. The rationalist has sought to demonstrate that simplicity entails

---

[105] For similar thoughts on the predicates of inner and outer sense, see R 4673, AA 17:368; R 5059, AA 18:75.

the imperishability of the soul as it is in itself. But regardless of the weakness of Kant's argument about incorruptibility, his insight that the compositeness the rationalist is concerned with when discussing the compositeness of the soul applies only to appearances and not to whatever noumenal substratum may ground the appearance of matter can be enlisted in order to provide a way of understanding how Kant might have concluded that Wolff was incorrect in thinking that a soul endowed with multiple powers would be a composite. As we have seen, Crusius was skeptical of Wolff's claim that a soul with multiple powers would be a composite of distinct substances that are the parts of the composite. One reason for this is that the Wolffian appears to think that anything that is a composite must be thought of as matter. This is to say that each substance is thought to occupy a distinct region of space and to be united with other substances that likewise occupy a distinct region of space. A composite thing has such spatial substances as its parts. Crusius rejects this view and maintains that a substance with multiple powers need not be regarded as a composite in the same way that matter is a composite.

Kant's distinction between appearances and things in themselves offers an interesting way to build upon Crusius's recognition that a substance with multiple powers need not be a composite in the sense that matter is a composite. By showing that the kind of compositeness that is associated with matter applies only to appearances and not things in themselves, Kant is able to explain why the soul, as the noumenal ground of thought, may not be treated mereologically as a composite of substances that occupy a distinct spatial region.[106] As we have seen, Kant maintains that only appearances may be regarded as a composite in this way; only they are extended and occupy a distinct region of space. The properties associated with matter such as compositeness, extension, and spatial location are properties only of outer sense and so do not apply to the substratum of thought but only to the manner in which we are affected by this substratum. This suggests that the powers with which the substratum that grounds thought is endowed may not be treated mereologically as parts of a whole substance each of which occupies some region in space. And the fact that the soul as the substratum of thought may not be thought of as composite matter would allow Kant to argue that the ground of the capacities for thought is simple although it potentially possesses multiple fundamental powers. It can possess these multiple powers without being a composite in the sense that

---

[106] In this vein, Kant also notes: "it is already in itself an unsuitable question to ask whether or not it [the soul] is of the same species as matter (which is not a thing in itself at all, but only a species of representations in us); for it is already self-evident that a thing in itself is of another nature than the determinations that merely constitute its state" (A 360).

matter is a composite. And Kant may argue this even while accepting the Wolffian premise that a composite must be thought of in terms of matter as an aggregate of spatially distinct parts.

Although Kant may align himself with Crusius to some degree in arguing that the fact that the soul has multiple powers does not entail that it is composite, this does not mean that he accepts the idea that the simplicity or non-compositeness of the soul entails anything about its immortality. We have seen that Kant argues that the soul may be composite and therefore corruptible as appearances although it is simple as a thing in itself. However, the fact that this argument is inadequate may have led to Kant's revision of his argument against imperishability in the B edition of the Paralogisms. In the B edition, Kant focuses on Mendelssohn's argument for imperishability. Mendelssohn was aware that the argument for imperishability from dissolution was still subject to the objection that the soul could nevertheless perish through annihilation. Nevertheless, Mendelssohn proposes that since a soul has no parts it cannot be diminished and gradually transformed into nothing since "there would be no time at all between a moment in which it is and another moment in which it is not, which is impossible" (B 414). Kant argues, however, that although a soul has no extensive magnitude, it nevertheless has intensive magnitude, i.e. "a degree of reality in regard to its faculties" (B 414) and so "could be transformed into nothing, although not by disintegration, but by a gradual remission (*remissio*) of all its powers (hence, if I may be allowed to use this expression, through elanguescence)." (B 414). As Kant notes, this is true of consciousness, and therefore of the soul as it appears in inner sense. More importantly, with regard to the previous discussion, this is also the case with the soul understood as a thing in itself that grounds thought through its powers. The powers of the soul could perish through elanguescence. So, whether the soul possesses a single power or whether it possesses a plurality of powers, neither scenario, even with respect to things in themselves, is sufficient to establish immortality according to Kant. In effect, Kant shows that immortality cannot be established on a theoretical basis even by engaging with the rationalist on his own terms by employing a priori arguments regarding the nature of things as they are in themselves.[107]

Having seen how Kant may argue against the thesis that a soul with multiple powers would be a composite, and having seen that he rejects the rationalist arguments that attempt to show that the simplicity of the soul entails its incorruptibility, we may return now to the

---

[107] On the method of engaging with the rationalists on their own terms, see the note beginning at B 415.

discussion of Kant's views on whether the substance that grounds thoughts has multiple powers. Although Kant does not offer a positive resolution to the question of how many powers ground thought outside of the statements in the subjective deduction, he does offer some clue in the Appendix to the Transcendental Dialectic, "On the regulative use of the ideas of pure reason" about the role of a fundamental power in metaphysics.[108] In the Appendix, Kant reiterates the findings of the Transcendental Deduction, namely that we appear to ourselves to have a variety of faculties – sensibility, consciousness, imagination, memory, wit, the ability to distinguish, desire and so on – although it is possible that these faculties may be grounded in a smaller number of faculties such as those identified in the Deduction as sensibility, understanding, and reason (B 676f.). However, the idea of a fundamental power that grounds all of these faculties is an idea of reason, which demands absolute totality in the synthesis of conditions and therefore also the reduction of all conditioned attributes to a single unconditioned substance. Although Kant does not say this, in a substance endowed with multiple powers, these powers would presumably be mutually conditioning insofar as they work together to endow the substance with the capacities it has and therefore also the unity of thought that arises from these capacities. But if such powers are mutually conditioning, then reason may still demand that we go further in our pursuit of an unconditioned ground of conditioned attributes. As Kant writes: "The idea of a fundamental power [*Grundkraft*] – though logic does not at all ascertain whether there is such a thing – is at least the problem set by a systematic representation of the manifoldness of powers" (A 649/ B 677). We proceed by comparing properties of powers and capacities in order to find what they have in common guided by the idea that there is a common power as their ground until we "bring them close to a single radical, i.e. absolutely fundamental, power" (A 649/ B 677). Although we can provide no a posteriori or a priori arguments establishing the existence of such a fundamental power, we may, however, use the idea as a means of organizing our investigation of mental faculties.[109]

---

[108] For Kant's additional discussions of the number of powers and our knowledge of fundamental powers, see: Metaphysik L_1 AA 28:262, AA 28:431, AA 28:432, AA 29:770, and R 4825, AA 17:739. Kant does sometimes seem to believe that the soul can have only one *Grundkraft*. See Metaphysik L_1 (AA 28:210, AA 28:261). Kant also sometimes appears suspicious of Crusius' proliferation of the powers of the soul, as in the Logik Blomberg (AA 24:82).

[109] Ameriks also discusses the influence of Crusius on Kant's thinking in the Appendix. See Karl Ameriks, *Kant's Theory of Mind. An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000), p. 246. As Ameriks points out, in the

We also began the discussion of Kant's views on the powers of the soul by indicating that the subjective deduction provides an argument for the irreducibility of sensibility and understanding to a fundamental power but does not provide an argument for how a soul may possess multiple powers without being a composite and fails to explain why the existence or non-existence of a fundamental power should remain unknown to us. We have seen how Kant may use the distinction between appearances and things in themselves to argue that the soul as the substratum of thought may possess multiple powers without being a composite in the sense that matter is a composite. The first answer also suggests an answer to the second question. As we have seen, Kant argues that the soul, as an unconditioned ground of thought, must be a thing in itself. Since we cannot cognize things in themselves directly, we remain ignorant of whether they possess a single fundamental power or multiple powers. Thus we must be satisfied with the analysis of our mental powers provided in the subjective deduction, although there may nevertheless remain some mystery about the number of powers of the soul as a thing in itself.

2.4 Conclusion

We began this chapter by considering Kant's response to debates among the German rationalists regarding the number of powers of the soul. It was shown that Wolff and Wolffian philosophers argue that the soul must be thought of as a simple substance that possesses a single power of representation that makes the unity of thought possible. The Wolffians also argue that a soul that possesses multiple powers would be a composite and so incapable of thought and as a composite would also be susceptible to perishing through a dissolution of its parts. Crusius and Lange, however, reject the Wolffian thesis that every power must be grounded in an independent substance and therefore also reject the idea that a substance endowed with multiple powers would be a composite of substance parts. They maintain instead that the soul is a simple substance endowed with multiple powers. Because the soul is simple it is also incorruptible. Kant takes up the discussion of the number of mental powers and their possible ground in a fundamental power in the subjective deduction,

---

Metaphysical Foundations of Natural Science, Kant actually posits two fundamental forces or powers of nature, attraction and repulsion. But this does not preclude the fact that mentality may be reducible to a single fundamental power.

arguing that thinking, or the unity of thought, is possible only if we have certain irreducible and jointly necessary mental powers. Kant's first pass at an answer to the question about multiple mental powers in the subjective deduction, however, leaves open a number of questions that might be raised by the Wolffian regarding whether Kant's view unacceptably entails that the soul is composite. We have also seen that Kant's answer to these concerns can be found in the Second Paralogism. Here Kant argues that there are neither a posteriori nor a priori reasons for thinking that the unity of thought cannot be grounded in a composite. Furthermore, we have seen that Kant rejects the idea that if the soul were endowed with multiple powers it would be a composite like matter. Kant has the resources to argue on the basis of his transcendental idealism that compositeness and spatial parthood apply only to appearances and not to things in themselves. So the soul as the noumenal ground of the unity of thought may possess multiple powers but not be a composite in the sense that worries the Wolffian. Kant also argues that regardless of whether the soul is simple and possesses a single power or multiple powers, the simplicity of the soul does not entail its incorruptibility because the soul could perish through a remission of its powers.

Having seen Kant's discussion of the powers that ground the unity of thought and his rejection of immortality, we may turn now to a consideration of a third crucial aspect in Kant's confrontation with the rationalist conception of the mind. One central reason the rationalist is concerned with establishing the simplicity and immortality of the soul is in order to argue that the soul may be susceptible to rewards and punishment in the afterlife. Having rejected the arguments for immortality, Kant nevertheless offers a positive view of the conditions under which actions may be imputable to the subject of thought. In chapter three, we will consider Kant's discussion of rationalist and empiricist views of personhood and show how Kant argues that certain mental powers are required to sustain personhood.

Chapter 3

The Metaphysics of Personhood in Kant's Third Paralogism

3.1 Introduction

In the previous chapters, we considered Kant's views on the role that substances and powers play in grounding thought and how he develops these views in contrast with and through a criticism of his predecessors, particularly Baumgarten, Wolff, and Crusius. We saw that Kant maintains that on the basis of the principle of sufficient reason one arrives at the idea that thought is an attribute of an absolute or ultimate substance endowed with a power whereby it grounds its attributes. We have also seen that Kant argues against the idea that the unity of thought must be the result of a single power of the soul, a *vis repraesentativa*. Instead, Kant argues that a number of mental powers are jointly necessary for the unity of thought and that the existence of multiple powers in the substance that grounds thought does not entail that this substance is a composite. We have also seen that Kant rejects the rationalist claim that the soul is a persisting, spatiotemporal, immortal substance. The issue of the immortality of the soul for the rationalists, however, concerned much more than the question of whether a simple soul could survive because it would not be subject to a dissolution of its parts. The rationalists were also concerned to show that the soul could have some memory of itself in the afterlife and thus that its personhood could be retained in the afterlife. Having rejected the implications of the substantial view of the self for any proof of the immortality of the soul, in the Third Paralogism Kant nevertheless considers what personhood consists in and what its role is in understanding the conditions under which previous actions may justifiably be imputed to someone. In this chapter, we will consider how Kant develops a positive view of personhood that is necessary and sufficient for moral responsibility.

In the Third Paralogism of the A edition of the *Critique of Pure Reason*, Kant presents the following argument as representative of rationalist arguments regarding the soul:

> What is conscious of the numerical identity of its Self in different times, is to that extent a person. Now the soul [is conscious of the numerical identity of its Self in different times]. Thus the soul is a person. (A 361)

Kant argues that rationalist philosophers have mistakenly argued from the major premise that one is a person insofar as one is "conscious of the numerical identity of its self in different times" to the conclusion that the soul is a person or retains personhood on the basis of a minor premise that asserts that the soul is "conscious of the numerical identity of its self in different times." According to Kant, however, the rationalist's conclusion regarding the persistence of the soul is an invalid "paralogism" because it is guilty of committing a *sophisma figurae dictionis*, or fallacy of the ambiguous middle term, regarding what it means for someone to be "conscious of the numerical identity of its self in different times."[110] Kant's diagnosis is that the rationalist has conflated the acceptable doctrine of the transcendental unity of apperception, which states that the "I think" must be able to accompany all of our representations in order for us to have coherent experience, with a substantial soul's empirical consciousness of itself. Although it is widely agreed that this captures Kant's critique of the rationalist, interpreters have often taken the aim of the Third Paralogism to be exhausted in this negative critique of rational psychology, and so have failed to see Kant as offering a positive view of personhood, or what he alternately calls personality or personal identity. One reason interpreters have tended to overlook Kant's positive contribution to metaphysical debates about personhood is that they tend to focus on Kant's epistemological critique of the rationalist So they have concerned themselves, for example, with whether Kant continues to believe as he did in the pre-critical period that we have an immediate consciousness of our identity, or whether and how we can be conscious of immaterial substances such as the soul given that the application conditions for the concept "substance" are lacking. The question for such interpretations is whether and how the second premise can be subsumed under the first and therefore whether the rationalist's conclusion follows. Without a doubt, a great deal, if not the majority, of Kant's discussion in the Third Paralogism revolves around this point. However, in contrast with these other interpretations, I suggest that we can gain a new insight into Kant's aims and the continuity of the discussion

---

[110] In the absence of the ambiguous middle term, the argument is valid and follows the rule *major sit universalis*, *minor affirmans* (The major should be universal, the minor affirmative); see Dohna-Wundlacken Logic, pp. 773–776. There is little consensus among interpreters on the details of how the rationalist's argument fails. For some of the different positions on this point, see C. Thomas Powell, *Kant's Theory of Self-Consciousness* (Oxford: Oxford University Press, 1990), pp. 130–135; James Van Cleve, *Problems from Kant* (New York: Oxford University Press, 1999), pp. 180–182; Karl Ameriks, *Kant's Theory of Mind*: *An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000), pp. 130–137.

of the Third Paralogism with other aspects of Kant's theoretical and moral philosophy as well as with the views of his rationalist and empiricist predecessors if we focus our attention on Kant's understanding of the first premise of the argument and its definition of personhood.[111]

In contrast with Patricia Kitcher who argues that the immediate historical context of Kant's account of personhood is Hume's empiricist account of the self as a bundle of representations, I argue in 3.2 that the most immediate historical context of Kant's discussion is the Lockean view of personhood and the response to this view among German rationalist philosophers, most notably Wolff and Leibniz.[112] The discussion of personhood with which Kant was familiar both from his readings of Locke and from his familiarity with the philosophy of Leibniz and Wolff is not primarily concerned with the epistemology of personal identity as in Hume's account of our inability to locate a substantial self, nor is it concerned primarily with the distinction between human persons and animals as one commentator has recently argued.[113] Rather, the main concern is with understanding the necessary conditions that must obtain for one to count as morally responsible for one's actions, with understanding the relationship between our identity as persons and our identity as moral subjects in juridical contexts and in the afterlife. Although commentators do not often acknowledge it, this conception of personhood is of central importance for the metaphysical foundations not only of Kant's theoretical philosophy but also for his practical philosophy since it concerns the conditions under which our free actions can be imputed to us. I show that the philosophical disagreement between Locke and the Leibnizian and Wolffian rationalists to which Kant is responding is not about whether consciousness of the identity of one's self in different times is necessary for personhood, a definition which all parties accept, but about the role substances play in grounding one's capacity for consciousness of one's identity and ensuring both the veracity of our consciousness of the identity of ourselves and ensuring that we may be held morally responsible even in cases in which the memory of our previous actions are not present to mind.

---

[111] I disagree in this regard with Van Cleve's claim that the third paralogism argument can be rewritten such that it does not include reference to persons but only to the endurance of the soul. On my interpretation, personhood, not merely endurance, is the central issue of the paralogism. See James Van Cleve, *Problems from Kant* (New York: Oxford University Press, 1999), pp. 180–182.

[112] See Patricia Kitcher, "Kant on Self-Identity," *Philosophical Review* 91(1) (1982), pp. 41–72.

[113] See Corey W. Dyck, "The Aeneas Argument: Personality and Immortality in Kant's Third Paralogism," *Kant Yearbook* 2 (2010), pp. 95–122.

In 3.3, I show that the importance of Kant's discussion of personhood in the Third Paralogism and the Transcendental Deduction of the A edition of the *Critique of Pure Reason* is that it is able to provide a conception of personhood that overcomes some of the epistemological problems with Locke's account of personhood without accepting the rationalist position that we have an immediate consciousness of ourselves as a substance that exercises the powers that ground consciousness. Kant overcomes the epistemic limitations of Locke's view by arguing that the consciousness we have of ourselves in different times that constitutes our personhood rests on certain mental capacities and powers that must operate conjointly in order to ensure that we have a coherent experience of ourselves and our world. The virtue of Kant's view is to show that we retain personhood not only when we are actually conscious of our identity but when we retain the mental powers that ensure that it is at least possible for us to become conscious of our identity in different times. In this regard, this interpretation differs from that provided by Karl Ameriks, which maintains that the Third Paralogism is about numerical identity over time and not the related question of "whether one has certain appropriate complex powers (at any time)."[114] For Kant, our personhood and our ability to exercise certain mental powers are inextricably intertwined such that our retention of these powers is necessary for the kind of consciousness that is required for personhood and so also for our moral responsibility for our free actions. In 3.4, I conclude by summing up the argument and pointing out conditions in addition to personhood that Kant maintains are necessary for moral responsibility each of which will be considered in the subsequent chapters.

3.2 The Lockean and Leibnizian Conceptions of Personhood

As I have suggested, when Locke, Leibniz, and the Leibnizian philosophers such as Wolff and Baumgarten with which Kant was familiar discuss the question of personhood, personality or personal identity, it is not primarily an epistemological question about whether and how one could know oneself to be a person who persists across time but is rather a question about the most appropriate conception of personhood that would allow one to make sense of the moral responsibility one has for one's actions. When Locke discusses the concept of a "Person," for example, he suggests that it is a "Forensick Term, appropriating

---

[114] See Karl Ameriks, *Kant's Theory of Mind*: *An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000), p. 129.

Actions and their Merit" and that "personality extends it *self* beyond present Existence to what is past […] whereby it becomes concerned and accountable; owns and imputes to it *self* past Actions."[115] Likewise for Leibniz, the question of personhood is intimately tied to moral identity. God's wisdom makes it such that the retention of personhood is necessary for us to be "sensitive to punishments and rewards."[116] Nor does Kant depart from this conception of personhood in the Anthropologie-Collins and his lectures on metaphysics when he writes: "personality makes it that something can be imputed [*imputirt*] to me" (Anthropologie-Collins, AA 25:11) and "This [consciousness of one's self] is psychological personality, to the extent they can say: I am. It further follows that such beings have *freedom*, and everything can be imputed to them; and this is *practical personality*, which has consequences in morality" (Metaphysik L_1, AA 28:277).[117] Personality, or personhood, was intended to explain the conditions under which our actions may be imputed to us and so also the conditions under which we could be morally responsible and justly rewarded or punished for our actions. And the issue of our moral responsibility for our actions is not only a practical issue in a forensic legal context in which it must be established whether one is responsible for a crime committed but is also an issue of religious importance concerning reward and punishment in the afterlife for our actions. In both cases, it is only if one retains one's personhood across the stretch of time in question that one's actions may justly be imputed to one and one can be susceptible to punishment or reward for these actions.

The definition of personhood expressed in the first premise of Kant's third paralogism, namely that "what is conscious of the numerical identity of its Self in different times, is to that extent a person," was common to empiricists and rationalists alike in the early-modern debates on personhood. Kant's formulation resembles Locke's view that personhood consists in the capacity of a thinking being to consider "it self as it self, the same thinking thing in different times and places."[118] Likewise, the German rationalist Christian Wolff's definition of personhood in his *Detusche Metaphysik*, with which Kant was familiar,

---

[115] See John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon Press, 1975), II.xxvii.26.
[116] G.W. Leibniz, *New Essays on Human Understanding*, ed. Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1996), II.xxvii.9.
[117] On Kant's distinction between psychological and practical personality, see Heiner Klemme, *Kants Philosophie des Subjekts* (Hamburg: Felix Meiner Verlag, 1996), pp. 97–101. See also: *Metaphysics of Morals*, AA 6:233; R 5646, AA 18:295; Anthropologie-Pillau, p. 2.
[118] Locke, John. *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon Press, 1975), II.xxvii.9–11.

states that "a thing is called a person that is conscious that it is the very same thing that was previously in this or that state."[119] Similarly, another German philosopher Johann August Eberhard, writes in his *Universal Theory of Thinking and Sensing* (1776): "the conservation of the I and of personhood depends simply on the consciousness of its uninterrupted persistence."[120] Each of these philosophers maintains some version of a consciousness-based account of personhood according to which personhood consists in our consciousness now of ourselves as the same person who undertook some previous action in the past or will suffer some future fate. Kant also makes clear in Metaphysik Dohna and elsewhere that our moral personality, that is to say the imputability of our free actions to us, is dependent upon our psychological personality, which is our ability to be conscious of previous actions.[121] This is to say that in legal contexts it would not be appropriate to hold someone responsible for some action if they could not at least recognize recall this action. And it would equally be contrary to God's wisdom to reward or punish us for previous actions if we could not at least be conscious of these actions.[122] Beyond their agreement about the definition of personhood, Locke and the rationalists also agree that certain mental powers are necessary for us to exercise the kind of consciousness required for personhood. Thus Wolff argues in this

---

[119] See Christian Wolff, *Verünftige Gedancken von Gott, der Welt under der Seele des Menschen, auch allen Dingen überhaupt* (*Deutsche Metaphysik*) (1720) (Halle: 1751), §924. Wolff also provides a memory criterion of personhood in his *Psychologia rationalis* as a "being which preserves a memory of itself, that is, which remembers that it is that same being that was previously in this or that state (*Persona* dicitur ens, quod memoriam sui conservat, hoc est, meminit, se esse idem illud ens, quod ante in hoc vel isto suit statu. Dicitur etiam *Individum morale*). See Wolff, *Psycholgia rationalis* (Frankfurt: 1734), §741.

[120] See Johann August Eberhard, *Allgemeine Theorie des Denkens und Empfindens* (Berlin: 1776), p. 25. See also Eric Watkins, *Kant's Critique of Pure Reason: Background Source Materials* (Cambridge: Cambridge University Press, 2009), p. 326.

[121] See Metaphysik Dohna (AA 28:683). Kant presents similar views of the relationship between consciousness and personhood at Metaphysik L_I, AA 28:296; Metaphysik Mrongovius, AA 29:911, 913; Metaphysik Volckmann, AA 28:411; Metaphysik Dohna, AA 28:680, 683, 688; Metaphysik K_2, AA 28:763; Metaphysik Vigilantius (K_3), AA 29:1036.

[122] Regarding the doctrine that we shall receive punishments and rewards in the afterlife for our actions in this life, Locke writes: "The Sentence shall be justified by the consciousness all Persons shall have, that they *themselves*, in what Bodies soever they appear, or what Substances soever that consciousness adheres to, are the *same*, that committed those Actions and deserve Punishment for them." See John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch, Oxford: Clarendon Press, 1975), II.xxvii.26. It might also be noted that being conscious of one's previous states as one's own may be necessary but not sufficient for regarding the action represented by this state as one's own. One might for example know that one was involved in some action but deny that this action was one's own and thus that one is morally responsible for it since one might deny that this action was done freely.

Deutsche Metaphysik (1720) that we are capable of being conscious of ourselves when we have the ability to distinguish ourselves from other things (§730). And this ability to distinguish ourselves from other things and thus to become conscious of ourselves requires that we can retain thoughts through time, distinguish them from each other, and recognize them as the same or different, all of which requires memory and reflection: "thus memory and reflection [*Überdenken*] produce consciousness" (§735).[123] Likewise, Locke argues that "the Mind has a Power, in many cases, to revive Perceptions, which it has once had" and beyond this capacity for memory also has other mental capacities, including a capacity for reflection, which are necessary for the kind of consciousness required for personhood.[124]

Despite this overall agreement on the definition of personhood and the consciousness required for it, they disagree, however, about the role the powers of substances play in grounding our consciousness. In his discussion of personhood, Locke insists that he does not wish to "meddle with any Physical Consideration of the Mind; or trouble […] to examine, wherein its essence consists, or by what Motions of our Spirits, or Alterations of our bodies, we come to have any Sensation by our Organs, or any Ideas in our Understandings; and whether those Ideas do in their Formation, any, or all of them, depend on Matter, or no."[125] Although he recognizes the need for certain mental capacities and even concedes that these capacities are grounded in the real constitutions and powers of substances, he is not concerned with explaining the relationship between these capacities and the real constitution of minds and the fundamental material or immaterial substances and powers that ground these mental capacities. Indeed, he ultimately argues that the real constitution of substances that ground our mental capacities is irrelevant for personhood. He writes: "For the same consciousness being preserv'd, whether in the same or different Substances, the personal Identity is preserv'd" by which he means that personal identity is independent of the

---

[123] In *Psychologia empirica*, Wolff defines the soul as "that being in us which is conscious of itself and of other things outside us" (Ens istud, quod in nobis sibi sui & aliarum rerum extra nos conscium est, Anima dicitur); see Wolff, *Psychologia empirica* (Frankfurt: 1737), §20.

[124] See John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon Press, 1975), II.x.2. This is not to say that Locke and Wolff think of mental powers such as memory or reflection in the same way.

[125] John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon Press, 1975), I.i.2. Lisa Downing suggests that for Locke "real constitution" is "the configuration of intrinsic and irreducible qualities responsible for all of a thing's qualities/powers." See Lisa Downing, "Locke's Ontology," in *The Cambridge Companion to Locke's Essay*, ed. Lex Newman (Cambridge: Cambridge University Press, 2007), pp. 370f.

persistence of whatever substances and powers consciousness is grounded in.[126] This position on the unimportance of substances and their causal powers in grounding personhood is motivated in part by Locke's epistemology, which claims that we can experience only the secondary qualities associated with consciousness and perception and cannot have any experience of the primary qualities or powers that ground these secondary qualities.[127] So Locke's claim of ignorance regarding substances leads him to focus solely on the role of consciousness in establishing personhood. In contrast, Christian Wolff argues that "the senses (§220), the imagination (§235), memory (§249), the faculty of reflection (§272), the understanding (§277), sensuous desires (§434), [and] the will (§492)" must be grounded in a simple substance and the single representative power of this simple substance.[128] Ultimately this disagreement about the role of substances and causal powers leads to a disagreement regarding some counterfactual possibilities for personhood between adherents of the Lockean and rationalist positions. Whereas Locke maintains that our consciousness and mental

---

[126] John Locke, *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon Press, 1975), II.xxvii.13.

[127] For Locke on the substances underlying consciousness, see Shelley Weinberg, "The Metaphysical Fact of Consciousness in Locke's Theory of Personal Identity," *Journal of the History of Philosophy* 50(3) (2012), p. 387; Margaret Atherton, "Locke's Theory of Personal Identity," *Midwest Studies in Philosophy* 8 (1) (1983), pp. 287–9; J.L. Mackie, *Problems from Locke* (Oxford: Oxford University Press, 1976/2005), p. 200f; Edwin McCann, "Locke on Identity: Matter, Life, and Consciousness," *Archiv für Geschichte der Philosophie* 69(1) (1987), pp. 75–76; Kenneth Winkler, "Locke on Personal Identity," *Journal of the History of Philosophy* 29(2) (1991), pp. 201–226.

[128] See Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (Halle: 1751), §747. The idea that Wolff like Leibniz and Locke is primarily concerned with the conditions under which one can count as morally responsible for one's previous actions in his account of personhood contrasts with the account provided by Corey W. Dyck who argues personhood among the German rationalists has primarily to do with the distinction between animals and human persons. However, although it is true that the rationalists and Wolff in particular argue that personhood is what distinguishes humans from animals, they are nevertheless concerned with the conception of personhood insofar as it can explain moral responsibility. The fact that animals lack personhood is one reason they are not considered morally responsible for their actions. See Corey W. Dyck, "The Aeneas Argument: Personality and Immortality in Kant's Third Paralogism," *Kant Yearbook* 2 (2010), pp. 95–122. In his account of the difference between persons and animals, Wolff also discusses the various faculties that are required for personhood. On the difference between humans and animals and how their possession or lack of mental faculties determines their ability to be conscious of their identity in different times, i.e. their personhood, see Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (Halle: 1751), §892, §924; *Psychologia rationalis* (Frankfurt: 1734), §767. Most importantly, animals lack memory: "Quoniam illud demum est persona, quod memoriam sui conservat; *bruta personae non sunt*" (*Psychologia rationalis* (Frankfurt: 1734), §767).

capacities may be annexed to different substances, because he believes there is no necessary relationship between our mental capacities and the substances that ground these capacities, the Wolffians maintain that our mental capacities are inseparable from the substantial, simple, soul that grounds these capacities, and thus that consciousness cannot be annexed to different substances.[129]

Beyond these metaphysical questions regarding the role of substances and powers in personhood, the rationalists also argue that there are epistemological problems that arise when one accepts as Locke and others do that continuity of consciousness is necessary and sufficient for personhood and therefore that the preservation of the substance that causally grounds the mental capacities required for consciousness of the identity of one's self in different times is irrelevant to the preservation of personhood. As Leibniz points out in the *New Essays*, which was first published in 1765 and so made a contemporary impact on the discussion of personhood in the German context within which Kant was working, in response to Locke's proposal, although our moral identity requires consciousness of our past actions, it is unclear how this identity can be divorced from the real identity of the substance that grounds consciousness.[130] For Leibniz and other rationalists, the Lockean account is problematic for a number of reasons. We may have lapses of consciousness, in dreams or drunkenness for example, and so on the Lockean account we would not be responsible for our actions in these times. We may also form false beliefs about our past experiences in the sense that we may be conscious of ourselves as having undertaken some action despite the fact that this consciousness has been annexed to a different substance that was not involved in the

---

[129] In *Psychologia rationalis*, Wolff argues that a person is "a singular, living substance" that is capable of being conscious of its identity in different times. See Christian Wolff, *Psychologia rationalis* (Frankfurt: 1734), §741.

[130] Leibniz writes: "I also hold this opinion that consciousness or the sense of *I* proves moral or personal identity." But he adds in response to Locke: "You seem to hold, sir, that this apparent identity could be preserved in the absence of any real identity. Perhaps that could happen through God's absolute power; but I should have thought that, according to the order of things, an identity which is apparent to the person concerned – one who senses himself to be the same – presupposes a real identity obtaining through each immediate [temporal] transition accompanied by reflection, or by the same sense of *I*; because an intimate and immediate perception cannot be mistaken in the natural course of things." See G.W. Leibniz, *New Essays on Human Understanding*, ed. Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1996), II.xxvii.9. For a discussion of Leibniz on Locke's view of personal identity, see Edwin Curley, "Leibniz on Locke on Personal Identity," in *Leibniz: Critical and Interpretive Essays*, ed. Michael Hooker (Minneapolis: University of Minnesota Press, 1982), pp. 302–326; Margaret Wilson, "Leibniz: Self-Consciousness and Immortality in the Paris Notes and After," *Archiv für Geschichte der Philosophie* 58 (1976), pp. 335–52.

undertaking of the action.[131] By allowing that consciousness could be annexed to different substances, Locke allows our psychological identity and our real identity as substance to come apart. Although such lapses of consciousness or false memories may be mitigated by the testimony of others for practical purposes in juridical contexts, Leibniz argues that ultimately moral responsibility requires not only our psychological identity but also our real identity as a soul that exercises a fundamental power of representation. The existence of the soul ensures both our veridical epistemological consciousness of the identity of ourselves in different times and the metaphysical fact of our real identity in different times. When we are conscious of the identity of our self in different times, we are immediately conscious of the real identity of the substance that grounds this consciousness. As Leibniz writes, "an intimate and immediate perception cannot be mistaken in the natural course of things."[132] And when we are not conscious of the identity of ourselves in different times, there is nevertheless some fact of the matter about our real identity in different times. Ultimately our real identity and the consciousness we have of it ensure that it is possible for us to be justly rewarded and punished for our actions both in juridical contexts and in the afterlife.[133] As we will see, in his discussion of personhood in the Third Paralogism and the Transcendental Deduction Kant is concerned with developing a conception of personhood that is necessary and sufficient for moral responsibility and that can overcome the epistemological problems associated with Locke's conception while also respecting the idea that we cannot have immediate consciousness of the ultimate substance or thing in itself that grounds the mental capacities required for consciousness of the identity of one's self in different times.

---

[131] See G.W. Leibniz, *New Essays on Human Understanding*, ed. Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1996), II.xxvii.9.

[132] See G.W. Leibniz, *New Essays on Human Understanding*, ed. Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1996), II.xxvii.9. The Wolffians also argue that we have immediate cognition of substances, and more importantly, that we have an immediate cognition of ourselves as immaterial souls that exercise our mental powers. For a discussion of Wolff on personhood, see Udo Thiel, *The Early Modern Subject* (Oxford: Oxford University Press, 2011), pp. 312–314.

[133] Leibniz does not, however, think that continuity of substance is sufficient in absence of continuity of consciousness: "[I]t is memory or the knowledge of this 'I' which makes it capable of punishment and reward. Likewise, the immortality which is demanded in morals and religion does not consist merely in this perpetual subsistence which belongs to all substances, for without a memory of what one has been, there would be nothing desirable about it." See "Discourse on Metaphysics" (1686) in G.W. Leibniz, *Philosophical Papers and Letters*, second edition, ed. and trans. Leroy E. Loemaker (Dordrecht: Kluwer Academic Publishers, 1989), § 34.

3.3 Kant on the Metaphysics of Personhood

In the period preceding the publication of the *Critique of Pure Reason*, Kant was well aware of the Lockean conception of personhood and the reception of this view by the Wolffians and by Leibniz in the *New Essays* in 1765.[134] In the 1770's, Kant recognizes that the Scholastic conception of personhood, which maintains that personhood consists in the preservation of the soul, must be supplemented by the recognition that the preservation of an immaterial soul is not sufficient for personhood in the absence of a consciousness of one's previous states, but Kant does not go so far as to accept the Lockean view that such consciousness is necessary and sufficient for personhood.[135] Much like Leibniz, in the 1770's Kant holds that our consciousness of the I and our previous states is annexed to the powers of an immaterial substance that underlies it, and of which thoughts are mere accidents. In Metaphysik L_1, he is reported to have said: "the I, or the soul through which the I is expressed, is a substance" (AA 28:266).[136] And in the Anthropologie-Collins, he suggests that "the proper [*eigentliche*] I is something substantial, simple and persistent" (AA 25:13). Kant also goes so far as to claim that we have an immediate, albeit indeterminate, cognition of ourselves as an immaterial substance that exercises various powers by means of which we ground accidents.[137] Insofar as we are conscious of the identity of ourselves in different times, we are also conscious of the real identity of the immaterial substance that underlies this consciousness and so are not susceptible to the kinds of epistemic problems associated with the Lockean view of personhood.

What shifts between Kant's account in the 1770's and his position in the *Critique of Pure Reason* is his recognition that there are impediments to our immediate knowledge of the powers of substances and thus to our knowledge of the real identity of the ultimate substance

---

[134] Kant explicitly discusses Locke and his view of personal identity and substances in: Metaphysik Herder (1762–74), AA 28:43; Metaphysik Dohna (1792–93), AA 28:682; Metaphysik L_2, AA 28:563.

[135] In Metaphysik Mrongovius (1782–83), Kant also denies that the persistence of a soul in the absence of consciousness of one's previous states is sufficient for personhood (AA 29:913). See also Metaphysik L_1, AA 28:296.

[136] See also: Anthropologie-Collins (1772–73), AA 25:2f.; R 4234 (1769–70), AA 17:470. Heiner F. Klemme also discusses the concept of the "I" in the Anthropologie-Collins in *Kants Philosophie des Subjekts* (Hamburg: Felix Meiner Verlag, 1996), pp. 78–79.

[137] For a discussion of Kant's claim that we have immediate cognition of ourselves as a substance, see Julian Wuerth, "Kant's Immediatism–Pre-Critique" *Journal of the History of Philosophy* 44(4) (2006), p. 532.

that grounds consciousness and the mental powers associated with consciousness. This recognition of our epistemic limitations comes with the development of transcendental idealism and the distinction between appearances and things in themselves along with the recognition that we cannot have immediate cognition of the intrinsic powers that ground appearances. Thus Kant argues that we can have knowledge only of the spatial and temporal properties of appearances, but we cannot cognize the nonspatiotemporal substances or intrinsic properties that ground these appearances and make them possible. This is a general recognition of our epistemic limitations with regard to substantial grounds that Kant also extends to our cognition of ourselves as the immaterial substance that grounds the accidents of inner sense and consciousness. So we find Kant in the Third Paralogism suggesting that we should not be misled as the rationalists have been misled to think that our consciousness of the identity of our self across various states is the consciousness or direct cognition of the identity of an immaterial substance that underlies our consciousness. Kant expresses this point most clearly in the B edition of the *Critique of Pure Reason* when he writes:

> [T]his identity of the subject, of which I can become conscious in every representation, does not concern the intuition of it, through which it is given as object, and thus cannot signify the identity of the person, by which would be understood the consciousness of the identity of its own substance as a thinking being in all changes of state […]. (B 408)

Kant also makes this point throughout the A edition Paralogisms as well such as when he writes: "The identity of the consciousness of Myself in different times is therefore only a formal condition of my thoughts and their connection, but it does not prove at all the numerical identity of my subject […]" (A 363).

Although Kant disputes our epistemic access to the substantial grounds of our conscious mental states, he nevertheless does not dispute the first premise of the rationalist's argument and the commonly accepted position that personhood consists in the consciousness of the identity of our self in different times.[138] And indeed he even goes so far as to say that this conception is "necessary and sufficient for practical use" (A 365f.). However, he also

---

[138] Kant's actual formulation is "What is conscious of the numerical identity of its Self in different times, is to that extent a person" (A 361), which suggests that Kant may allow for degrees of personhood, as might be attributed to children, the disabled, or even to some types of animals. As I will suggest, Kant means to say that we are persons to the extent that we retain certain mental capacities that make it possible for us to become conscious of the identity of our self in different times. This formulation also allows for the possibility that when our mental capacities, such as memory or the understanding, are deficient, then we have a lesser degree of personhood and are therefore less morally responsible for our actions.

notes that this conception of personhood "can remain" only "insofar as it is merely transcendental, i.e. a unity of the subject which is otherwise unknown to us, but in whose determinations there is a thoroughgoing connection of apperception" (A 365f.). With this caveat, Kant invokes the doctrine of the transcendental unity of apperception from the Transcendental Deduction, which maintains that in order for us to have coherent experience our representations must be attributed to a single subject, or in other words, that the "I think" must be able to accompany all our representations.

Kant's allusion to apperception is, however, far more complex and interesting in the context of the Third Paralogism if one considers the relationship of apperception to other mental faculties in the Transcendental Deduction. In the A edition of the Transcendental Deduction, from which Kant is drawing in his discussion in the A edition Paralogisms, apperception is one mental faculty among others that contribute to the possibility of our having a unified and coherent experience of our selves and of the surrounding world. According to Kant:

> There are, however, three original sources (capacities or faculties of the soul), which contain the conditions of the possibility of all experience, and cannot themselves be derived from any other faculty of the mind, namely sense, imagination, and apperception. (A 94)[139]

These faculties include sense, imagination, and apperception, faculties that make various other mental syntheses possible whereby particular sensations are united into larger representations of objects, brought under concepts, and attributed to a single self. Among the kinds of synthesis performed by the mind which contribute to an overall unified consciousness, Kant identifies the 'synthesis of apprehension in intuition', which involves the synthesis of appearances of perceived objects, the 'synthesis of reproduction in the imagination', and the 'synthesis of recognition in a concept'. The second operation of synthesis, 'synthesis of reproduction in the imagination', in particular plays a central role in facilitating the continuity of mental life because it serves the basic function of representing objects that are not themselves the object of immediate apprehension in sensibility and so serves as a kind of memory insofar as it reproduces previously apprehended representations. This kind of faculty is necessary for consciousness of one's identity in different times

---

[139] In contrast with commentators such as Strawson who believe one can separate the analytical portions of Kant's discussion in the Deduction from his commitment to faculty psychology, we can see that faculty psychology is central to Kant's account of personhood. See P.F. Strawson, *The Bounds of Sense* (London: Methuen, 1966).

because in order to be conscious now of some previous representation one must be able to recall it. The details of how these various faculties operate together in order to generate unified conscious experience across time are complex, but it is sufficient to note that not only is apperception needed for a coherent and unified consciousness, each of these irreducible faculties and powers of synthesis is necessary for the possibility of the kind of unified conscious experience Kant attributes to us.[140] In this regard, Kant is similar to both Locke and Wolff in recognizing that the unity and continuity of our mental life involves a variety of mental capacities.[141]

In his discussion of synthesis and our mental capacities in the Transcendental Deduction, Kant is also explicit that we may sometimes be aware of ourselves as carrying out the synthesis of our representations and of exercising the kinds of capacities that allow us to synthesize our representations. This has led a number of commentators to associate our awareness of our numerical identity, our consciousness of the identity of ourselves in different times, with our awareness of our synthesis of representations or our capacity for synthesizing. Kant suggests such a view in the B edition when he writes that apperception "contains a synthesis of the representations, and is possible only through the consciousness of this synthesis" (B 133). Such passages have led Patricia Kitcher to argue that conscious of the identity of one's self for Kant means explicit or implicit consciousness of the synthesis of representations in rational cognition. She writes for example: "As we know from the B deduction, a subject can be conscious of her identity *only* through being conscious of adding different representations together (B 133–34)."[142] According to Kitcher's view of synthesis, synthesis is the process whereby intuitions are subsumed under judgments and judgments are

---

[140] Although in the B edition, Kant abandons the "subjective deduction" of the A edition, which attempted to explain "the objective validity of *a priori* concepts" on the basis of the cognitive faculties involved in cognition (B xvi) and instead focuses on the "objective deduction," he nevertheless maintains that the mental faculties of sensibility, understanding, and apperception are required for coherent experience in both editions.

[141] In contrast, however, with Wolff and other Leibnizians who argue that our mental faculties are reducible to a single faculty of representation, Kant argues in the subjective deduction that we have three mental irreducible and jointly necessary mental faculties: sensibility, understanding, and apperception. For a discussion of the subjective deduction, see Corey W. Dyck, "The Subjective Deduction and the Search for a Fundamental Force," *Kant-Studien* 99(2) (2008), pp. 152–179.

[142] See Patricia Kitcher, *Kant's Thinker* (Oxford: Oxford University Press, 2011), p. 191. See also Patricia Kitcher, "Kant's Paralogisms," *Philosophical Review* 91(4) (1982), p. 536. On counting, see Patricia Kitcher, *Kant's Thinker* (Oxford: Oxford University Press, 2011), pp. 128, 146.

related to one another in rational cognition. For example, when we count, we see our thought of a number "two" as dependent on an antecedent thought "one" and so see these representations as rationally connected or synthesized.[143] And one is conscious of the identity of one's self at different times when one is explicitly or implicitly conscious of synthesizing representations and judgments in such rational cognition and of attributing them to the same "I" in transcendental apperception.[144] Similarly, Eric Watkins has argued on the basis of the B 133 passage that representations can be attributed to oneself only if one is aware of the activity of connecting these representations together. This means we can have an indirect awareness of our identity by being "aware of the activity of the self when it connects its various representations and by then inferring that it is one and the same self that does the connecting."[145]

---

[143] Kitcher holds that such consciousness is necessary for personhood but rejects the idea that it is sufficient (although she does concede that it is necessary and sufficient for practical use). On Kitcher's interpretation of the Third Paralogism, Kant could not have held that consciousness is sufficient for personhood because this would rule out the possibility of religious conversion, which requires that one become a new person. She writes: "Since Kant wants to make room for conversion, he has to dismiss not just substantial identity, but also memory continuity and sameness [of] cognitive subject—the logical I—as sufficient for sameness of person." See *Kant's Thinker* (Oxford: Oxford University Press, 2011), p. 186. On conversion and becoming a "new man," Immanuel Kant, *Religion within the Boundaries of Mere Reason*, trans. and ed. Allen Wood and George di Giovanni (Cambridge: Cambridge University Press, 1998) AA 6:73f.

[144] This conception of rational connectedness also accords with earlier work in which she argues Kant maintains that there is an existential dependence of mental states on one another. Mental states are existentially dependent upon one another if and only if an antecedent mental state is necessary and sufficient for a subsequent mental state, and such mental states are necessary and sufficient for one another when they are synthesized in rational cognition. See Patricia Kitcher, "Kant on Self-Identity," *Philosophical Review* 91(1) (1982), pp. 41–72. On Kitcher's view, Kant introduces the existential dependence of mental states on one another in response to Hume's denial that there is an identical self to be found in consciousness. Against Hume, Kant can maintain that one is conscious of the identity of one's self in different times because one is conscious of the existential dependence of representations upon one another and is aware of oneself as the one who connects or synthesizes these representations. As I have argued, however, Kant is primarily concerned with the view of personhood proposed by Locke and the rationalists rather than with Hume's denial of knowledge of a substantial soul. Although it might be argued on Kitcher's behalf that existential dependence among representations also allows her to maintain that personhood is retained even in lapses of consciousness, since these connections would hold even in the absence of consciousness of these connections, she appears to maintain that consciousness of one's identity requires existential dependence and awareness of this existential dependence, i.e. awareness of oneself as synthesizing representations.

[145] See Eric Watkins, "Kant's Model of Causality: Causal Powers, Laws, and Kant's Reply to Hume," *Journal of the History of Philosophy* 42(4) (2004), pp. 480.

However, both of these interpretations of the role synthesis plays in personhood will result in a problem for Kant's account. If our awareness of our activity of synthesizing representations is necessary for consciousness of our identity and so also for personhood, then we cease to be persons in a variety of circumstances such as daydreaming, drunkenness, and other moments in which we are not wholly concentrated on our mental activities. By requiring awareness of our activity of synthesizing representations, such interpretations would fall prey to the objection raised against Locke's view of personhood that lapses of consciousness would entail a break in personhood. The problems related to lapses of consciousness are especially exacerbated by Kitcher's interpretation of synthesis, which requires the rational synthesis of representations in judgments in order to retain personhood. This understanding of synthesis would entail that we fail to retain personhood across any period when we are not reasoning such as when we are dreaming, sleeping, or simply confused. But Kant certainly allows for the synthesis of sensible impressions in such a way that these representations are not incorporated into rational cognition.[146]And Kant also recognizes that although synthesis is an "indispensable function of the soul," it is one "of which we are seldom ever conscious" (A 78/B 103), so it would be surprising if Kant required that such an awareness of synthesis is necessary in order to retain personhood since we seldom ever have such awareness.

Such problems, however, can be overcome if we understand that Kant's requirements for personhood are much less strict and that he allows that we can retain personhood even in cases in which we are not directly aware of our synthesizing representations into a unified consciousness. According to Kant, for representations to count as one's own, and therefore for them to be incorporated into the continuity of one's mental life, one need not actually be aware of the activity of synthesizing representations into a continuity. Rather, it must only be *possible* to synthesize one's representations and become aware of them as a unity so

---

[146] Karl Ameriks calls this a "pre-judgmental unity." See "Kantian Apperception and the Non-Cartesian Subject," in *Kant and the Historical Turn: Philosophy as Critical Interpretation* (Oxford: Oxford University Press, 2006), p. 55. See also the discussion of obscure representations in the *Anthropology* (AA 7:135), and the discussion of the subjective empirical unity of consciousness that occurs through association of representations (B 139–40). Kant also allows that all representations are temporally ordered in time without the use of the categories; see *Anthropology* (AA 7:161) and the *Critique of Pure Reason* (B 51). He also allows that continuity of consciousness can occur even with obscure representations at R 4562 (1772–1776), AA 17:594.

synthesized. Kant puts this requirement most clearly in the B edition Deduction, when he writes:[147]

> The I think must *be able to* accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is to say that the representation would either be impossible or else at least would be nothing for me. (B 131–132)

> The thought that these representations given in intuition all together belong to me means, accordingly, the same as that I unite them in self-consciousness, or at least *can* unite them therein, and although it is itself not yet the consciousness of the synthesis of these representations, it still presupposes the possibility of the latter, i.e., only because I can comprehend their manifold in a consciousness do I call them all together my representations; for otherwise I would have as multicolored, diverse a self as I have representations of which I am conscious. (B 134)

But it is also evident in the A edition Transcendental Deduction that Kant states a rather modally-weak requirement that it must only be *possible* for representations to by synthesized into a unity and for one to become aware of the unity so synthesized in order for these representations to be part of one's continuous consciousness. One need not be explicitly conscious of the activity of connecting representations, nor need one be explicitly aware of one's previous states as one's own in order for them to be representations that *possibly* belong to one. While in a fugue state, for example, my mental faculty of sensibility operates insofar as I am taking in sensations of the world, though I may not unite these experiences into a coherent experience. Nevertheless all of these representations are possible objects of my unified experience insofar as they may later be taken up in explicit cognitions and synthesized into a coherent experience. In this regard, we can see that Kant maintains that we retain our personhood so long as it is *possible* to synthesize representations into a coherent unity of consciousness and in this way to become conscious of our numerical identity across different states. And it is possible to synthesize representations in this way so long as we retain the mental capacities or powers Kant identifies as sensation, imagination or understanding, and apperception.

In the preceding, I have argued that in the Third Paralogism Kant accepts the definition of personhood as consciousness of the identity of one's self in different times but rejects the idea that consciousness of the identity of one's self in different times entails the real identity of one's self as soul or substance. Similarly to Locke, Kant appears to accept the consciousness-based account because he maintains that we cannot have knowledge of the

---

[147] I have added italics to emphasize the modal strength of Kant's requirement.

mind-independent substances or things in themselves that ground consciousness and thought. Since we have no epistemic access to things as they are in themselves, we have no reason to think that consciousness of the identity of one's self in different times implies the real identity of one's substance much less consciousness of the real identity of one's substance. However, consciousness-based accounts are also susceptible to the problem that lapses of consciousness entail a lapse of personhood, something that the rationalist accounts overcome by appealing to the real identity of substance. We have seen that Kant may overcome this problem by suggesting that personhood is retained so long as one retains the mental capacities required for one to synthesize one's representations into a continuous consciousness. The debate regarding the role of consciousness and the real identity of our self as substance was, however, also concerned with another issue that may now be addressed. We have seen that because Locke believes that personhood consists in continuity of consciousness he may also argue that consciousness may be annexed to different substances. Rationalists such as Wolff and Leibniz, however, deny this possibility since they maintain that consciousness of the identity of one's self in different times entails the real identity of one's substance. I now turn to a consideration of Kant's position on whether consciousness or mental capacities could be annexed to a different substance and whether, if they could, personhood would be preserved.

There is a great deal of evidence that Kant was interested in this question, particularly in the A edition of the *Critique of Pure Reason*. In his discussion in the A Deduction of the "three original sources" "which contain the conditions of the possibility of all experience" Kant explicitly refers to these as "capacities or faculties of the *soul*" (*Fähigkeiten oder Vermögen der Seele*) (A 94). And in A 78/ B 103 when Kant discusses synthesis as the effect of the imagination, he refers to synthesis as "a blind though indispensable function of the *soul*" (*Funktion der Seele*). Capacities and consciousness were thus considered to be attributes of a substantial soul. Kant was, however, ambivalent about the commitment to a soul that grounds mental attributes as is evidenced by the fact that the paragraph at A 94 containing the reference to the soul is excised from the B edition, and in his personal copy of the A edition, Kant replaces the reference to the "function of the *soul*" (*Funktion der Seele*) at A 78/ B 103 with "function of the understanding" (*Funktion des Verstandes*) (AA 23:45, E24).[148] Kant's interest in the A Deduction with the soul as the ground of our mental capacities also carries over into his explicit discussions in the A edition Paralogisms of the

[148] I have added italics to these passages for emphasis.

unknowable though necessary ground of appearances. In the second Paralogism, for example, he argues that the mental capacities required for unifying representations may be grounded in a composite of substances. Kant does also revise the reference to the soul as the ground of mental capacities in the B edition Paralogisms just as he does in the B edition Transcendental Deduction in favor of formulations that do not carry with them a commitment to the existence of a substantial thing in itself that grounds these capacities.[149] Nevertheless, despite his subsequent revisions, Kant's discussions in the A edition suggest that he was initially thinking of mental capacities as grounded in the substance that underlies consciousness just as Locke and the Wolffians did.

In the A edition of the Third Paralogism, Kant explicitly takes up the problem that faced the Lockeans and Wolffians. Using the analogy of a billiard ball, he considers whether consciousness may be annexed to another substance and whether personhood may be preserved in such a scenario. He writes:

> An elastic ball that strikes another one in a straight line communicates to the latter its whole motion, hence its whole state (if one looks only at their positions in space). Now assuming substances, on the analogy with such bodies, in which representations, together with consciousness of them, flow from one to another, a whole series of these substances may be thought, of which the first would communicate its state, together with its consciousness, to the second, which would communicate its own state, together with that of the previous substance, to a third substance, and this in turn would share the states of all previous ones, together with their consciousness and its own. The last substance would thus be conscious of all the states of all the previously altered substances as its own states, because these states would have been carried over to it, together with the consciousness of them; and in spite of this it would not have been the very same person in all these states. (A 363–4)

In the scenario envisioned here, consciousness is transferred from one substance to another such that only the final substance in the series retains consciousness of all of its previous states. This substance would be conscious of all of the previous states although its consciousness was annexed to different substances at different times.

There are two related questions that may be raised in interpreting Kant's position on this problem. The first is whether he believes it is conceivable that consciousness or the mental capacities that make consciousness possible could be transferred from one substance to another. And the second is whether he believes that personhood would be retained in such a scenario. Julian Wuerth has argued that Kant denies that consciousness could be transferred

---

[149] Rolf-Peter Horstmann also notices such a change between the two editions of the Paralogisms; see "Kants Paralogismen," *Kant-Studien* 84(4) (1993), pp. 408–425.

form one substance to another because Kant thinks that even God could not place the accident of one substance into another substance. This claim is supported primarily by passages in which Kant argues against the view that causation occurs through a transfer of accidents from one substance to another and maintains instead that substances influence one another by causing reciprocal changes in the states of the substances.[150] It is important to recognize, however, that the question Locke and others are raising is not whether a token-identical property could be transferred from one substance to another through God's will or through some other mechanical process but whether two different substances may instantiate type-identical properties. Although there may be doubts about whether Kant believes God could transfer a token-identical property, or attribute, from one substance to another, there is no reason to think that he would deny that two substances could have type-identical properties. In other words, although he may deny that property P could be token identical in substances A and B, he would appear to allow that property P could be type identical in substances A and B. The scenario envisioned does not require the actual causal transfer of a token-identical property from one substance to another but only requires that a type of property can be instantiated by more than one substance.[151] If Kant were to deny that two substances could instantiate type-identical properties, he would be committed to denying, for example, that two different people could each have blue eyes. In Metaphysik Herder, Kant

---

[150] According to Wuerth "[i]t makes no sense to speak of divorcing a mode from its substance as though it were a separate existence; for Kant, thoughts are not like books on a bookshelf that can conceivably be transferred to another bookshelf." See Julian Wuerth, "Kant's Immediatism–Pre-Critique," *Journal of the History of Philosophy* 44(4) (2006), p. 529f. Wuerth refers to R 3783 (1764–68), AA 17:292; Metaphysik Mrongovius, AA 29:770; Anthropologie-Collins (1772–73), AA 25:15.

[151] This is also how Locke understood the matter. Regarding whether consciousness of past actions could be transferred from one thinking substance to another he writes: "I grant, were the same Consciousness the same individual Action, it could not: But it being a present representation of a past Action, why it may not be possible, that that may be represented to the Mind to have been which really never was, will remain to be shown." See *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Clarendon Press, 1975), II.xxvii.13. However, this claim also leads Reid to object to Locke's account, and one might also think that Kant's account is susceptible to this objection. According to Reid, "when Mr. Locke therefore speaks of "the same consciousness being continued through a succession of different substances" […] these expressions are unintelligible to me unless he means not the same individual consciousness, but a consciousness that is similar or of the same kind" (175). But if the transferred consciousness is merely type identical, then it seems that persons A and B are not identical but merely similar. See Thomas Reid, "Of Mr. Locke's Account of Our Personal Identity" Essay III, chapter VI of *Essays on the Intellectual Powers of Man* (1785); reprinted in *Essays on The Powers of the Human Mind* (London: 1827), p. 175.

also provides a clue to the conditions under which a transfer of type-identical properties between substances might be possible when he discusses the rationalist doctrine that "each subject in which an accident inheres must itself contain a ground of its inherence" (AA 28:52). On this view, which Kant appears to endorse, God could not produce an accident in a substance through only his own external power. Rather, the substance itself must also have the power to ground the inherence of this property. As Kant quips, "otherwise I could also produce thoughts in a mere wooden post, if it were possible by a mere external power."[152] This means that in order for consciousness to be annexed to another substance, the substance itself would have to have the power required to ground the inherence of the attribute of consciousness. This discussion suggests that Kant thinks that it is conceivable that consciousness, and the various mental capacities that make consciousness possible, could be annexed to a different substance so long as this consciousness is only type identical and the substance to which the consciousness is annexed retains the power necessary to support such consciousness.

As I presented it above, consciousness might be understood as a memory of some particular action or state. Interpreted this way, the question is whether, for example, my memory of being in Berlin could be transferred to a different substance. However, for Kant the issue may be slightly more complicated. When speaking of the possible transfer of consciousness from one substance to another in Kant, we are discussing not only the possibility of transferring a set of memories, but also of transferring the transcendental unity of apperception. Understood in this way, the question is whether my unity of apperception, or as Kant sometimes says, the "I think," can be transferred from one substance to another. It appears on the basis of what was argued previously that a token-identical "I think" could not be transferred from one substance to another but a type-identical "I think" could. But this does not quite capture the scenario envisioned, since in wondering whether the "I think" could be transferred from one substance to another, one is not merely wondering about whether some general type of "I think" could be transferred, but whether someone's own particular consciousness or apperception expressed in the "I think" could be transferred. But although the issue appears in this sense to be more complicated, it should not be. It appears plausible to think that the unity of apperception of one person may be transferred to another substance as long as the actual property is not being transferred from one substance to another but only an instance is transferred. The scenario only appears more complicated when

[152] See Metaphysik Herder, AA 28:52.

one equates apperception with the "I think." It appears more complicated because of the indexical "I." We generally think that the "I" refers to whatever it is annexed to, so it seems perplexing to think that an "I" that refers to substance A could be transferred to substance B and yet remain the same "I" that refers to substance A. The confusion, however, appears to arise only form the semantics of the indexical "I." And this complication can be avoided by simply formulating the question as one of whether consciousness as the unity of apperception can be transferred from one substance to another. And we have seen that Kant thinks that it is at least conceivable that it can.

But even if Kant concedes that such a scenario is possible, does he believe that personhood would be retained in such a scenario? A clue to Kant's position on this question can be found when he claims that despite the fact that the last substance would retain consciousness of the states of the previous substances "it would not have been the very same person in all these states" ("dem unerachtet, würde sie doch nicht eben dieselbe Person in allen diesen Zuständen gewesen sein") (A 363–4). There are a few ways to understand this passage. On one reading, the passage would appear to suggest that Kant is denying that personhood would be retained in the Lockean scenario. He does not, however, provide any compelling reasons for why one should think this is the case. And indeed, his denial that personhood is retained seems counterintuitive because he appears to be siding with the rationalist who claims that the real identity of the substance underlying consciousness is necessary for personhood, which is in tension with his previous claims that consciousness of the identity of one's self in different times does not entail the real identity of the ground of consciousness. Perhaps this tension can be resolved if we recognize that although Kant is clear in the surrounding discussion that consciousness of the identity of one's self in different times does not entail that one can have knowledge of the real identity of the substance that underlies our conscious mental states, this does not mean that he does not believe there may be other independent grounds for accepting that personhood requires the real identity of the substance that supports consciousness as an accident. However, on another reading of this passage, Kant means to say that consciousness would be transferred from one substance to another despite the fact that the person, qua substantial soul on the rationalist conception, would not have been the same in each state. Thus Kant is merely restating the scenario in a way that alludes to the rationalist view that personhood requires a real soul substance. Although the soul substance might be different in each case, consciousness is transferred. Now, if Kant does in fact hold that the possible consciousness of the identity of one's self in different times is necessary and sufficient for personhood, then it seems that Kant may be

interpreted as claiming along with Locke that personhood may be preserved in a case where consciousness is annexed to a different substance or soul. This latter reading is also consistent with Kant's idea that it is possible that type-identical consciousness can be transferred from one substance to another so long as the substances have the power required to sustain consciousness.

In contrast with Locke, however, who dismisses speculation about the substance or substances that ground consciousness, Kant recognizes that the substance that grounds consciousness plays an important role in personhood even if the persistence, i.e. real identity, of the substance that grounds consciousness is not necessary for personhood. According to the rationalist, personhood requires that the substance that grounds consciousness is a persisting substance. One reason the rationalist adopts such a view is because a persisting entity would be able to sustain the possibility of a continuing consciousness even in lapses of consciousness and thus overcome a major problem with the Lockean consciousness-based account of personhood. Regarding the role such a substance plays in sustaining the possibility of consciousness, Kant writes:

> It is remarkable, however, that personality, and its presupposition, persistence, hence the substantiality of the soul, must be proved only now for the first time. For if we could presuppose these, then what would of course follow is not the continuous duration of consciousness, but rather the possibility of a continuing consciousness in an abiding subject, which is already sufficient for personality, since that does not cease at once just because its effect is perhaps interrupted for a time. (A 365)

According to Kant, if we could presuppose that a persisting substance grounds consciousness, then it would follow not that there is a continuing consciousness but the *possibility of a continuing consciousness in an abiding subject* (A 365). The possibility of continuing consciousness in an abiding subject would also be sufficient for personhood because the possibility of a continuing consciousness does not cease "just because its [the substance's] effect is perhaps interrupted for a time" (A 365). What Kant means by this is that even if consciousness is interrupted for some period of time, the possibility of a continuing consciousness, and thus personhood, is retained because the substance that grounds the possibility of a continuing consciousness persists, i.e. has real identity across time. Although consciousness as an effect of this persisting substance may be interrupted for a period of time, due to sleep or drunkenness for example, the substance that grounds this effect nevertheless continues to exist and thus sustains the possibility of continuing consciousness. Importantly, however, in providing such an account, the rationalist must assume that a real substance persists across time and sustains the possibility of continuing consciousness. Kant is,

however, clear throughout the Paralogisms that the rationalist makes use of an illicit concept of substance when he argues that the unknowable substance that grounds consciousness is a persisting spatiotemporal entity. According to Kant, the rationalist view is misguided because we cannot apply the concept of substance as a persisting entity to the thing in itself that grounds consciousness. So, Kant would reject the idea that we should assume that such a persisting substance grounds the possibility of continuing consciousness.

Nevertheless, Kant does appear to believe that a substance properly understood has an important role to play in sustaining the possibility of a continuing consciousness and therefore also personhood. According to Kant, we may understand a substance not as a persisting entity but as merely a ground of attributes, which grounds attributes through a power. The existence of such a substance or substances as a ground of consciousness and mental capacities is also necessary and sufficient for personhood.[153] It is necessary and sufficient for personhood because this substance through its powers sustains the possibility of a continuing consciousness by sustaining the various mental powers that Kant argues work together in order to produce a unified conscious experience through the synthesis of representations. In contrast with the rationalist view of personhood, which requires the real identity of a persisting substance, Kant only requires a substance that can sustain certain powers, but it need not be a persisting substance. Indeed, Kant allows for the possibility that a series or multiplicity of substances could ground the possibility of continuing consciousness. This conception of personhood also allows Kant to respond to worries about lapses in consciousness. On the interpretation I have proposed, Kant may argue that so long as there is a substance that grounds the various mental capacities needed for possible consciousness of the identity of one's self in different times, one retains one's personhood. In addition to providing a response to the problem of lapses in consciousness, this view of personhood may also provide a response to the problem of false memories that was raised for the Lockean account by Leibniz. Recall that Leibniz objects that Locke's theory would appear to allow that one might merely seem to remember some previous action although one was not in fact the person who undertook this action. On Leibniz's view, this problem is resolved by arguing

---

[153] According to Kant, the idea that a substance grounds inner attributes is a legitimate postulate of pure reason in its heuristic use as a guide to empirical scientific investigation. Kant writes: "Following the ideas named above as principles, we will first connect all appearances, actions, and receptivity of our mind to the guiding thread of inner experience as if the mind were a simple substance that (at least in this life) persists in existence with personal identity, while its states – to which the states of the body belong only as external conditions – are continuously changing" (A 672/ B 700).

that when we are conscious of the identity of our self in different times, we are conscious of the real identity of our subject. Kant's view, however, suggests that there is a mind-independent fact of the matter about whether one retains one's personhood or not independent of what one may believe about one's past states and actions and whether they belong to one. One retains personhood so long as the substance or substances that ground the faculties required for the possibility of one's continuing consciousness retain the power to ground these faculties. Although one may be mistaken about one's past, there is nevertheless an independent fact of the matter about whether the substance that has the power to ground one's ability to remember and be conscious of previous states and actions retains this power. So long as it does, one has the possibility of a continuing consciousness even if one happens to have a foggy or errant memory about the details of past events and actions.[154]

3.4 Conclusion

In the preceding discussion, I have argued that Kant's discussion of personhood in the Third Paralogism is not exhausted by his criticism that the rationalist oversteps the legitimate boundaries of our capacity for knowledge by claiming that consciousness of the identity of one's self in different times entails the real identity of a substantial soul. Much like Locke and the rationalists, Kant also provides a positive conception of the kind of personhood that is necessary for moral responsibility. According to the positive account I have attributed to Kant, he agrees with both his empiricist and rationalist predecessors that a continuing consciousness of the identity of one's self in different times is necessary for personhood. For Kant this means that personhood requires the retention of certain mental capacities or powers – sense, understanding, apperception, and reason among others – that make it possible for one

[154] We may illustrate the view with an example. Imagine that we wonder whether x at t_1 and y at t_2 are the same person. Kant's view states that they are if and only if some substance supports the mental capacities required for y at t_2 to be possibly conscious of some of the states of x at t_1. Consider the scenario where someone is in an accident that results in amnesia. The Lockean view says that x at t_1 (pre-amnesia) and y at t_2 (post-amnesia) are not the same person because y at t_2 has no consciousness of x at t_1. The view I have attributed to Kant holds that y at t_2 is the same person as x at t_1 person if and only if y at t_2 is possibly conscious of x at t_1. And y at t_2 is possibly conscious of x at t_1 just in case certain mental faculties are retained, mental faculties that in turn require certain substances as their grounds. It remains open to some degree what such substances are. One might, for example, think that personhood is retained just in case one retains certain parts of the brain that are necessary for synthesizing one's representations. Or such substances may be things in themselves.

to synthesize representations into a coherent and continuous experience.[155] Kant's conception of personhood also allows him to explain whether personhood is retained if consciousness is annexed to different substances, how personhood is retained across lapses of consciousness, and how there is a fact of the matter about personhood independent of how one appears to oneself in consciousness.

In addition to providing an account of personhood required for morally responsibility, Kant's account of the role of mental capacities in personhood also reveals a great deal about the kind of personhood required for moral agency. The various faculties Kant identifies as allowing for the synthesis of representations in continuing consciousness are the same mental capacities that are employed in cognition and in moral deliberation.[156] In order to deliberate about some action, for example, we receive information about the situation through our senses,and we make judgments about the situation through the understanding. The conception of personhood that Kant offers is therefore not merely of theoretical interest and only tangentially related to his practical philosophy. A person in Kant's sense is one who retains the faculties required for consciousness of the identity of one's self, which are the same capacities required for moral agency. In this regard our psychological personhood provides the necessary conditions for our moral personhood, i.e. our susceptibility to reward and punishment but also our practical agency and ability for ethical deliberation. In this regard,

---

[155] Colin Marshall has also recently argued for a metaphysical view of personal identity in Kant. On Marshall's account, Kant holds an "effect-relative view" of the self according to which "for any particular unified experience, whatever thing or things are immediately causally responsible for the unity of that experience compose a self" and personal identity consists in the identity of this self. See Colin Marshall, "Kant's Metaphysics of the Self," *Philosophers' Imprint* 10(8) (2010), p. 16. Although my account of personhood in Kant also shows that certain capacities and substances are necessary and sufficient for the unity of our mental life, my interpretation differs in a very important respect insofar as it does not maintain that an actual unified conscious experience is needed for personhood but only the possibility of a unified conscious experience for which our mental capacities and the substances that underlie them are responsible.

[156] Christine Korsgaard has argued against Parfit's Lockean conception of personhood that although one might grant Parfit's metaphysical claims regarding persons, from a practical standpoint we must recognize that persons exhibit a certain unity over times as agents: "from a moral point of view it is important not to reduce agency to a mere form of experience. It is important because our conception of what a person is depends in a deep way on our conception of ourselves as agents." She finds such an agent-centered view of morality in Kant. See Christine M. Korsgaard, "Personal Identity and Unity of Agency," in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), p. 364. On my interpretation, Kant is able to retain agency in his account of personhood by making certain faculties necessary for both personhood and agency.

Kant presents a more developed understanding of the links between personhood and agency than that presented by Locke and other consciousness-based accounts of personhood.

We have also seen that the rationalist maintains that personhood requires the real persistence of a simple, immaterial substance or soul. Since it is incorruptible, such a soul persists in the afterlife, and because it is conscious of the identity of its self in different times, it may also be justly rewarded and punished for its actions. This general view, which is shared by all of the rationalists Kant considers, is also dualist insofar as it posits a substantial immaterial soul that inhabits a material body. Although the body can perish because it is composite and material, the soul persists because it is simple and immaterial. In the next chapter, we will consider Kant's criticism of substance dualism and the rationalist account of mind-body interaction and his views on the possibility of mind-body interaction in the Paralogisms of Pure Reason.

Chapter 4

Kant's Critical Solution to the Problem of Mind-Body Interaction

4.1 Introduction

In the previous chapter, we considered Kant's view on personhood and saw that for Kant the preservation of personhood is necessary for moral responsibility. According to Kant, personhood consists in the possible continuity of consciousness of one's self in different times. Someone is possibly conscious of the identity of themselves in different times if and only if they retain the mental capacities required for synthesizing representations into unified experience, which are grounded in the power of a substance. In this chapter, we will consider another aspect of Kant's discussion of rational psychology, namely the relationship between mind and body, Kant's reasons for rejecting the dualist picture provided by the rationalist, and his proposed critical solution to the problem of mind-body interaction.

In the *Critique of Pure Reason*, Kant argues that the problem of mind-body interaction, or the problem of the "community in which our thinking subject stands to things outside us" (A 389), which has troubled substance dualists such as Descartes and Wolff, can be resolved if we recognize that mental and physical substances are appearances and not things in themselves. Although most interpreters agree on the general critical framework of Kant's response to the dualist and its reliance on transcendental idealism, there is a great deal of disagreement about whether Kant merely criticizes the foundational commitments of substance dualism or whether he also offers a positive metaphysical explanation of the possibility of interaction in response to the dualist. Interpretations that are motivated by an epistemological understanding of Kant's transcendental idealism and its claim that we cannot know things as they are in themselves have often taken Kant merely to be asserting an *ignorabimus*, namely that mind-body interaction among things in themselves is a pseudo-problem since we cannot know whether and how such interaction occurs independent of appearances.[157] C. Thomas Powell, for example, argues that it is not "terribly surprising" that

[157] On Kant's *ignorabiums*, see Wilfrid Sellars, "…this I or He or It (The thing) which thinks…," *Proceedings and Addresses of the American Philosophical Association* 44 (1970/1971), p. 11.

Kant's discussion of mind-body interaction undergoes an elanguescence from his pre-critical work to the A edition of the *Critique of Pure Reason* and eventually the *Prolegomena* because "given Kant's adoption of the position of transcendental idealism, the mind/body problem is a problem about the interaction of noumena. And that is something about which Kant – at his best – has nothing to say."[158] This repeats a common view, particularly within a certain strand of interpretation that includes Strawson and Bennett, that Kant's positive pronouncements about things in themselves are best thought of as indiscretions to which he would not admit in the light of day.[159] However, such interpretations appear unsatisfactory given the central if unspoken role of noumenal mind-body interaction in Kant's theoretical and practical conception of the self and moral responsibility, which requires a positive explanation of the possibility of interaction rather than merely an assertion of our ignorance of how interaction is possible. In the absence of an account of Kant's positive views on the possibility of mind-body interaction, for example, his account of freedom of the will and its claim that we act freely as things in themselves appears to be an obscure anomaly that does not fit with Kant's denial of knowledge of things in themselves. Likewise, his claims regarding our moral responsibility for actions appears to require not only an understanding of whether we are free in our actions but also an understanding of how the mental causation required for moral action is possible in the first place.[160] Moreover, the issue of mind-body interaction is not relegated to the discussion in the Transcendental Dialectic but is also central to Kant's discussion of how the mind is affected by objects (A 19/ B 33) and his claim that the mind is affected not by empirical objects but by "something supersensible which underlies the former, and of which we can have no knowledge," i.e. by things in themselves

---

[158] See: C. Thomas Powell, "Kant's Fourth Paralogism," *Philosophy and Phenomenological Research* 48(3) (1988), p. 414; C. Thomas Powell, *Kant's Theory of Self-Consciousness* (Oxford: Oxford University Press, 1990), pp. 199–200.

[159] See: Jonathan Bennett, *Kant's Analytic* (Cambridge: Cambridge University Press, 1966); Bennett, *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974); P.F. Strawson, *The Bounds of Sense* (London: Methuen, 1966).

[160] Derk Pereboom also recognizes that Kant's views on transcendental freedom require an understanding of things in themselves in terms of causal powers; see "Kant on Transcendental Freedom," *Philosophy and Phenomenological Research* 73(3) (2006), pp. 544–6. Desmond Hogan also recognizes that Kant aims in part to establish the metaphysical preconditions of freedom and argues that Kant's theory of freedom provides the foundation for his argument for the indispensability of noumenal affection; see Desmond Hogan, "Noumenal Affection," *Philosophical Review* 118(4) (2009), p.532.

(AA 8:215).[161] Without an account of Kant's thoughts on the possibility of mind-body interaction, and in particular, the possibility of how the mind is affected or interacts with things in themselves, this central, if often disputed, doctrine is left underdeveloped.[162]

What is needed in grounding both Kant's view of the affection of the mind by things in themselves and his moral philosophy is an account of how and why Kant claims that mind-body interaction is possible or at least unproblematic on the basis of his ontological distinction between appearances and things in themselves. Although Karl Ameriks and others have made some progress in this regard by recognizing that Kant claims that mental and physical appearances interact in virtue of the things in themselves that ground them, a more detailed interpretation of Kant's critical view on the possibility mind-body interaction requires a thorough understanding of the structure of this grounding relation and the powers through which things in themselves cause appearances.[163] The aim of this chapter is to offer such an interpretation. Section 4.2 presents Kant's critical response to substance dualism and argues for an interpretation that shows that Kant presents a positive response to the problem of the possibility of substantial mind-body interaction. Kant's proposed solution to the problem of the possibility of interaction follows upon his pre-critical interest in providing an account of substantial mind-body interaction and contrasts with Humean constant-conjunction and contemporary counterfactual accounts of causation. I also argue that materialist, pneumatist, and phenomenalist interpretations of Kant's proposed solution to the problem of interaction are unsatisfactory. Section 4.3 presents Kant's positive ontology of the

---

[161] Kant also equates this supersensible ground of appearances with things in themselves. See A 42/ B 59; A 387; A 494/B 522; A 537/B 565; A 566/ B 594; AA 4:289; AA 4:319; AA 4:451.

[162] F.H. Jacobi famously objects to Kant's claim that unknowable things in themselves affect the mind because it is inconsistent with Kant's doctrine of our ignorance of things in themselves. See F.H. Jacobi, *David Hume über den Glauben; oder Idealismus und Realismus* (Breslau: 1787). Desmond Hogan identifies four historical responses to this issue: A deflationary account; one that holds that Kant simply contradicts himself; one that maintains it is merely a personal conviction and so not in contradiction with the doctrine of ignorance; and one that maintains we are ignorant only of intrinsic properties and not the causal relations whereby they affect us. See Desmond Hogan, "Noumenal Affection," *Philosophical Review* 118(4) (2009), p. 502f. My aim in this paper is not to resolve Kant's apparently contradictory claims but only to show how such affection might be possible on a certain interpretation of Kant's transcendental idealism.

[163] See: Karl Ameriks, *Kant's Theory of Mind*: *An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000); Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005).

relationship between mental and physical appearances and things in themselves. I argue that Kant retains a form of neutral monism according to which mental and physical appearances are heterogeneous and grounded in but not reducible to things in themselves, which are homogenously neither mental nor physical. Kant's account of the relation between appearances and things in themselves and the reciprocal causal interaction of things in themselves in virtue of their powers allows him to explain how substantial mind-body interaction may be possible given the doctrine of transcendental idealism. Section 4.4 considers and replies to some objections to this interpretation of Kant. And section 4.5 concludes with a summary and points toward problems that will be considered in the next chapter.

4.2 Substance Dualism and Kant's Critical Solution to the Possibility of Mind-Body Interaction

In his discussion of mind-body interaction following the Fourth Paralogism in the A edition of the *Critique of Pure Reason*, Kant suggests that the general confusion that arises when philosophers are faced with explaining the interaction of heterogeneous mental and physical substances arises from a conflation of appearances and things in themselves. He writes:

> [A]ll the difficulties that concern the combination of thinking nature with matter arise without exception solely from the surreptitious dualistic notion that matter as such is not an appearance, i.e., a mere representation of the mind, which corresponds to an unknown object, but is rather an object in itself, as it exists outside us and independently of all sensibility. (A 391)

In order to overcome the problem of the dualism of mental and physical substances philosophers such as Descartes, for example, propose that the community between thinking and extended substances can be explained in terms of physical influx, while Malebranche proposes an occasionalist solution according to which God intervenes to ensure the interaction of substances, and Leibniz and Wolffian philosophers following Leibniz argue that mental and physical substances simply mirror one another in a universal harmony. According to Kant, however, these views, particularly the physical-influx view, are dubious because "the sort of community that is claimed to occur between two species of substances, thinking and extended, is grounded on a crude dualism," which takes physical substances to be things in themselves rather than appearances or "representations of the thinking subject"

(A 392).[164] The rationalists mistakenly take matter and material substance to be a thing in itself rather than a manner in which things appear to us in part because they do not recognize the distinction between appearances and things in themselves and so also fail to see that properties associated with matter and material bodies are properties that only spatial and temporal appearances may possess.[165]

Kant argues, however, that if one recognizes this distinction between appearances and things in themselves, then there is a critical solution to the problem of the possibility of mind-body interaction. In the B edition of the *Critique of Pure Reason* he writes:

> But if one considers that the two kinds of objects [mental and physical substances] are different not inwardly but only insofar as one of them appears outwardly to the other, hence that what grounds the appearance of matter as thing in itself might perhaps not be so different in kind, then this difficulty vanishes […]. (B 427–8)

Here Kant suggests that if we recognize that mental and physical substances are not so different in kind, then a simple and elegant solution to the problem of the possibility of interaction is at hand. There are a few ways to understand Kant's proposal for explaining the possibility of mind-body interaction. One way to read both the A 391 and B 427–8 passages above is as claiming that mental and physical substances are not so different in kind because a physical substance is merely an appearance or a representation in a mind. Because it is merely a representation in the mind, it is mental just like the mind. And since there is no real heterogeneity in kinds of substances, mind-body interaction is possible. This reading suggests that Kant may have been proposing a phenomenalist solution to the possibility of mind-body interaction. I will consider this interpretation in more detail throughout the chapter but would first like to propose and consider an alternative interpretation that appeals to some of Kant's additional statements on how the problem of the possibility of interaction may be resolved.

Another way to interpret Kant is as claiming that if we take transcendental idealism seriously, then we see that although mental and physical substances may be different in kind as appearances, "inwardly" as things in themselves they may not be so different in kind. This line of thinking is particularly evident in the A edition, when Kant writes: "The

---

[164] See A 389–391. Kant often uses the same German term *Gemeinschaft* for the Latin terms *communio* and *commercium* (A 213–4/ B 260–1). The former has the sense of "mutual participation," denoting association, while the latter has the sense of "trade," or "commerce," denoting interaction. See C. Thomas Powell, *Kant's Theory of Self-Consciousness* (Oxford: Oxford University Press, 1990), pp. 189–191.

[165] Although Kant mentions here only that the dualist takes matter to be a thing in itself rather than a mere mode of representing things, it is also clear as we will see that this criticism applies equally to how the dualist understands the mental as a thing in itself.

transcendental object that grounds both outer appearances and inner intuition is neither matter nor a thinking being in itself, but rather an unknown ground of those appearances that supply us with our empirical concepts of the former as well as the latter" (A 379f.). Similarly, in the Second Paralogism, he also writes: "that same Something that grounds outer appearances […] considered as noumenon (or better, as transcendental object) could also at the same time be the subject of thoughts […]" (A 358–9). The point Kant makes in these passages is that although appearances are heterogeneously mental or physical, there is no reason to think that the "substrate" or things in themselves that ground these appearances is heterogeneous with respect to this distinction. The same kind of thing in itself that grounds mental appearances may also ground physical appearances. And if it is true that the ground of appearances is homogeneous in this regard, then the problem of the heterogeneity of substances that motivates the problem of mind-body interaction "vanishes." On this interpretation, Kant demonstrates the possibility of interaction by simply removing the major hindrance to interaction, namely the heterogeneity of mental and physical substances.

Before developing this interpretation further by considering what Kant might mean by a ground of appearances that is homogenous with respect to the distinction between mental and physical and considering the alternative interpretations of Kant's proposal, it is important to note that Kant appears to agree with the dualist both that mental and physical substances are irreducibly heterogeneous in some regard and, more importantly, that a solution to the problem of the possibility of interaction requires an account of substantial causation according to which a cause in one substance produces or brings about an effect in another substance.[166] This is a position that Kant retains from his pre-critical attempts at a solution to the problem of interaction in the *Thoughts on the True Estimation of Living Forces* (1747) and the *New Elucidation of the First Principles of Metaphysical Cognition* (1755), where he offers various versions of a physical-influx solution to the problem of mind-body interaction. By retaining a commitment to an account of interaction that involves causation or interaction among substances, Kant implicitly rejects the Wolffian and Leibnizian doctrine of pre-established harmony as well as the occasionalism of Malebranche and others. Kant also forgoes some alternatives that might resolve the problem of interaction by showing that the problem of interaction arises because of the conception of causation rather than the heterogeneity of substances. On Hume's account of causality, for example, which holds that

---

[166] On mental causation and these varieties of causation, see Karen Bennett, "Mental Causation," *Philosophy Compass* 2/2 (2007), p. 319.

causation is simply constant conjunction, the dualist does not face any problem in explaining interaction. The dualist may simply argue that mental events follow physical events and vice versa in a reliable and predictable way without any need for an explanation of interaction. Or, on a counterfactual conception of causation, the dualist might argue that to say that an event E_1 causes an event E_2 simply means that if E_1 had not occurred E_2 would not have occurred. For example, to say that your interest in mind-body interaction caused you to read this chapter means that if you had not been interested in mind-body interaction, you would not have read this chapter. On either Hume's account or the counterfactual conception of causation, the causal interaction between mind and body is unproblematic. The fact that Kant does not offer a revision of the concept of causation is, however, not entirely surprising given his pre-critical views and the account of causation he provides in the Analogies of Experience, which requires the persistence and reciprocal interaction of substances.[167]

It is worth noting that the interpretation of Kant's response to the dualist being developed here, which maintains that Kant is proposing an account of the possibility of interaction that requires the substantial interaction of the things in themselves that ground mental and physical appearances, differs greatly from an interpretation that might rely on an epistemological account of Kant's distinction between appearances and things in themselves. On the epistemological account, Kant is not proposing that things in themselves are the grounds of appearances. Rather, the distinction between appearances and things in themselves is merely a distinction between two ways of considering an object, according to our spatial and temporal forms of intuition or independent of these forms of intuition.[168] In

---

[167] See: A 176/ B 218 – A 218/ B 265; AA 4:312–313. Kant also rejects Hume's account of causation in the *Prolegomena*, but he does not appear to do so by appealing to substantial interaction. Here he seems to accept the idea that cause and effect are merely correlated, although he argues that we may legitimately hold them to be necessarily correlated in contrast with Hume who holds that we cannot maintain that there is a necessary correlation between cause and effect. See AA 4:310–3. On Kant's rejection of Hume's account of causation, see Eric Watkins, "Kant's Model of Causality: Causal Powers, Laws, and Kant's Reply to Hume," *Journal of the History of Philosophy* 42(4), pp. 449–488.

[168] For a concise overview of contemporary interpretations of Kant's transcendental idealism, see Dennis Schulting, "Kant's Idealism: The Current Debate," in *Kant's Idealism: New Interpretations of a Controversial Doctrine*, ed. D. Schulting and J. Verburgt (Dordrecht: Springer, 2011), pp. 1–28. For representative epistemological interpretations of transcendental idealism, see: Henry Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, Revised & enlarged edition (New Haven: Yale University Press, 2004); Gerold Prauss, *Kant und das Problem der Dinge an Sich* (Bonn: Bouvier, 1974). For various versions of metaphysical interpretations of transcendental idealism, see: Erich Adickes, *Kant und das Ding an Sich* (Berlin: Pan Verlag Rolf Heisse, 1924); Lucy Allais,

order for objects to be objects of knowledge for us, they must have the spatial, temporal, and cognitive properties that they receive from our spatial and temporal forms of intuition and the categories. Because only appearances have such properties, only they can be objects of knowledge for us. However, we may also consider objects as they are in themselves in abstraction from these properties, but such objects cannot be objects of knowledge for us. It might be argued on the basis of such an interpretation of transcendental idealism that Kant's point about dualism is that considered according to the forms of intuition, appearances are mental and physical, but considered independent of such forms of intuition, things in themselves are homogeneously neither mental nor physical and so conceivably capable of interaction. Under one description that posits heterogeneous mental and physical properties, interaction is problematic, whereas under another description that does not posit such properties interaction is unproblematic. Although there may be textual support for such an interpretation, such a response to the dualist on Kant's part would not be philosophically satisfying because it merely suggests that whether interaction is problematic depends upon how the situation is described. But the question of whether you have enough money to pay the rent becomes no less problematic depending on whether you describe yourself as having enough money or not. It might also be argued on the basis of the epistemological interpretation that Kant's point is only that we should remain agnostic about whether and how mind-body interaction is possible since we cannot know anything about things as they are in themselves.[169] But it is unclear why our ignorance of whether things as they are in themselves are capable of interaction makes interaction unproblematic. A problem remains a problem even if we cannot know what the solution to it might be.

However, Kant's epistemic point that we cannot have knowledge of things as they are in themselves may also be given an interpretation that supports a stronger response to the dualist. Kant's epistemic point is not merely that we cannot know whether things in

---

"Kant's One World," *British Journal for the History of Philosophy* 12(4) (2004), pp. 655–68; Allais, "Kant's Idealism and the Secondary Quality Analogy," *Journal of the History of Philosophy* 45(3) (2007), pp. 459–484; Tobias Rosefeldt, "Dinge an sich und sekundäre Qualitäten," in *Kant in der Gegenwart*, ed. J. Stolzenburg (Berlin: de Gruyter, 2007), pp. 167–209; Rae Langton, *Kantian Humility* (Oxford: Oxford University Press, 1998).

[169] C. Thomas Powell argues that Kant holds that the heterogeneity of mental and physical appearances is a problem only for appearances and that since we have no way to decide whether this problem obtains at the level of things in themselves, interaction is not a problem for us; see *Kant's Theory of Self-Consciousness* (Oxford: Oxford University Press, 1990), pp. 193–6. In contrast, I argue that Kant is providing a metaphysical solution rather than merely arguing for epistemic agnosticism.

themselves are mental or physical but that for all we can possibly know, things in themselves are neither mental nor physical. Rather than providing a dogmatic objection to dualism which would require "an insight into the constitution of the nature of the object, in order to be able to assert the opposite of what the proposition claims about the object" (A 388), Kant's critical demonstration of the possibility of interaction does not require one to have a better acquaintance with the constitution of the nature of the object, i.e. things as they are in themselves. Rather, it only requires knowledge that only appearances are spatial and temporal and therefore also physical and mental. This point regarding things in themselves nevertheless puts Kant in a position to offer a critical demonstration of the possibility of interaction. Given our forms of intuition, appearances are mental and physical, whereas things in themselves are intrinsically neither mental nor physical. Kant makes this point regarding the things in themselves that ground appearances, when he writes, for example: "The transcendental object that grounds both outer appearances and inner intuition is neither matter nor a thinking being in itself, but rather an unknown ground of those appearances that supply us with our empirical concepts of the former as well as the latter" (A 379f.). We need not have direct acquaintance with things in themselves to know that they are neither mental nor physical because we can know that only appearances are mental or physical and things in themselves are not appearances. Although this answers to some degree the worry interpreters might have that Kant is overstepping his epistemic limitations by making positive claims about things in themselves in his proposed solution to the problem of the possibility of interaction, more still needs to be said about what it means to claim that things in themselves are neither mental nor physical and why the fact that they are neither mental nor physical allows for the possibility of their interaction.

Before doing so, it is worth mentioning why Kant may have rejected some alternative metaphysical responses to the dualist, which hold that the ground of appearances is homogeneous and so capable of interaction. It perhaps goes without saying that Kant rejects materialism or physicalism as a solution to the problem of interaction because he quite explicitly rejects materialism throughout his writings. There is also no clear reason why Kant would have enlisted the distinction between appearances and things in themselves in order to argue that materialism is true of things as they are in themselves instead of arguing straightforwardly for materialism. More importantly, however, materialism is at odds with Kant's critical solution to the problem of freedom of the will, which requires that we are free as things in themselves and subject to causal determinism as appearances. If things in themselves were homogeneously material, they would be subject to causal determinism and

incapable of freedom. Kant likewise rejects the construal of the ground of mental and physical appearances as immaterial in the sense that things in themselves might be thought of as immaterial penetrable pneumata that interact according to certain fixed laws. As Kant writes: "[I]f one wants to broaden the concept of dualism as it is usually applied and take it in a transcendental sense, then neither it, nor the pneumatism that is opposed to it on the one side, nor the materialism on the other side, have the least ground, since then one's concepts would lack determination, and one would take the difference in the mode of representing objects, which are unknown to us as to what they are in themselves, for a difference in these things themselves" (A 379).

An alternative possibility that has some support in Kant's explicit statements in the *Critique of Pure Reason* may be to interpret Kant as arguing for some form of phenomenalism. The support comes most explicitly in the passages I mentioned above, where Kant writes that "all the difficulties that concern the combination of thinking nature with matter arise without exception solely from the surreptitious dualistic notion that matter as such is not an appearance, i.e., a mere representation of the mind" (A 391), and "But if one considers that the two kinds of objects [mental and physical substances] are different not inwardly but only insofar as one of them appears outwardly to the other" (B 427–8). On one version of phenomenalism proposed by James Van Cleve, things in themselves are minds and physical objects are constructions in such minds.[170] Although Van Cleve does not indicate how this phenomenalist interpretation of Kant might be used to solve the problem of the possibility of mind-body interaction, it might be argued that if things in themselves are minds, and physical appearances are constructions in such minds, interaction would presumably be unproblematic because all interaction of physical and mental substances would be the representation of interaction among physical and mental substances or would be interactions among mental items. Whether one finds this alternative interpretation of Kant's account of the possibility of mind-body interaction plausible will likely depend on how plausible one finds it to interpret Kant in general as a phenomenalist. The one obvious reason to reject the phenomenalist interpretation is that Kant explicitly distinguishes his transcendental idealism from Berkeley's phenomenalism in the *Prolegomena to any Future Metaphysics* (1783), although some interpreters have argued that Kant maintains some form of phenomenalism that is distinct form Berkeley's phenomenalism. Another objection to the phenomenalist interpretation is that it simply cannot account for Kant's claim that we have

---

[170] See James Van Cleve, *Problems from Kant* (New York: Oxford University Press, 1999).

knowledge only of our representations and his claim that things in themselves exist although we cannot know them.[171] If our minds are things in themselves, and we have knowledge of our minds, then we have knowledge of things in themselves. Kant, however, claims that we cannot know our minds or mental states as they are in themselves.[172] These objections are not decisive but are intended only to indicate that there are reasons to seek an alternative approach that can avoid some of these objections while still accounting for Kant's positive response to the dualist. In what follows, I aim to explain the nature and consequences of Kant's claim that the possibility of mind-body interaction is unproblematic because things in themselves are intrinsically neither mental nor physical.

4.3 Kant's Neutral Monism and the Possibility of Mind-Body Interaction

4.3.1 Kant's Neutral Monism

In the foregoing discussion, I have followed Kant in equating the denial that things in themselves are temporal or spatial with the denial that they are mental or physical, thinking or extended. Although Kant does not provide an explicit argument for why these are equivalent, one might be reconstructed from his claim in the Transcendental Aesthetic that space and time are forms of intuition rather than properties of things as they are in themselves. According to Kant, only appearances and not things in themselves are in space and time. This is to say that spatial properties such as "being located to the right of x" or temporal properties such as "simultaneous with x" are properties that apply only to appearances insofar as they are given spatial and temporal form by our forms of intuition and do not apply to things as they are in themselves independent of these forms of intuition. These spatial and temporal forms of intuition are also necessary conditions of the experience of objects of inner and outer sense (A 49).[173] If we did not have a spatial form of intuition, objects of outer sense would not appear to us in space. And if we did not have a temporal form of intuition, objects of inner sense would not appear to us in time. This means that we can apply the properties associated with inner and outer sense only to those things that are subject to our spatial and temporal forms of intuition. Since things in themselves are not subject to the forms of

---

[171] See Kant, Bxxvi, A 251–2; *Prolegomena* AA 4:315. For a discussion of problems with the phenomenalist interpretation of transcendental idealism, see Lucy Allais, "Kant's One World," *British Journal for the History of Philosophy* 12(4) (2004), pp. 655–684.
[172] See A 38/ B 54–5, B 150–7.
[173] See also A 30/ B 45, A 105, A 252.

intuition, we cannot apply the properties associated with inner and outer sense to things in themselves. Kant associates the properties of inner and outer sense with mental and physical properties respectively. Thus the properties of inner sense include mental properties such as thinking and believing, and the properties of outer sense include physical properties, most notably extension and composition. If the above is correct, then when Kant says that neither spatial nor temporal properties apply to things in themselves, this also means that neither physical nor mental properties apply to things in themselves.

It is also important to recognize the scope of Kant's denial that things in themselves are mental and physical in light of his views on spatial and temporal properties. Kant is denying only that the temporal properties associated with the mental and the spatial properties associated with the physical apply to things in themselves. He is not suggesting that *no* mental or physical properties apply to things in themselves. An example may help to clarify what is at stake here. Kant argues that the mental property of thinking cannot be applied to things in themselves. What he means by this is that things in themselves do not have the property of thinking in the same sense that a person endowed with a temporal form of intuition does. What characterizes our particular kind of thinking is that it occurs in time. We reason in time from premises to conclusions, or we deliberate in time about what course of action to take. This is a temporal sense of thinking. Kant does not, however, deny that non-temporal properties associated with thought could be applied to things in themselves. One might think here of the medieval idea associated with Boethius that God's foreknowledge is timeless. God need not have knowledge of one event before having knowledge of its consequent. Although it is denied, for example, that God knows things in time, God may nevertheless know things. Likewise for Kant, the properties associated with the mental are not univocal. Mental and physical properties do not apply in the same way to appearances and to things in themselves. Spatial and temporal physical and mental properties apply to appearances, and their non-spatial and non-temporal counterparts may apply to things in themselves. Thus Kant may allow for the idea of a non-temporal choice in his account of freedom of the will.[174] And he also appears, I will argue, to allow for non-extended, non-

[174] For a discussion of the notion of timeless agency in Kant's account of freedom of the will, see Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 73–101.

spatial properties such as those associated with powers to apply to things in themselves in his account of how things in themselves interact and ground appearances.[175]

Kant's insight that mental and physical properties apply only to appearances is interesting for the period for a few reasons. Descartes and the Wolffian rationalists to whom Kant was responding in his discussion of dualism also construed the distinction between the mental and physical as one between things that had the essential property of thinking and those whose essential property was extension.[176] But Kant diverges from these philosophers and makes an important contribution to the debate about mental and physical properties for the period by arguing that the properties associated with thought and extension are dependent upon our spatial and temporal forms of intuition rather than upon anything essential about the substances in question. Indeed, Kant's recognition that mental and physical properties are dependent upon forms of intuition also allows him to show how the impasse regarding the interaction of mental and physical substances arises from the claim that mental and physical properties are properties of things as they are in themselves rather than things as we represent them to be. Thus although, it was widely suggested by Descartes' critics that mind-body interaction is problematic because of the heterogeneity of substances, Kant is able to provide an explanation of this heterogeneity that does not appeal to the essential properties of these substances and to show that they are heterogeneous only as they appear and not as they are in themselves.[177] Kant's insight is also independently interesting because it also shows that a solution to the problem of the possibility of interaction need not appeal to a revision of our concept of substantial causation, as Humeans or some contemporary philosophers might argue, but can retain a realist view of causation and appeal to a revision of our conception of mental and physical substances and their properties.

Having clarified the nature of Kant's claim that things in themselves are neither mental nor physical, more still needs to be said about the relationship between these things in themselves and mental and physical appearances. Kant argues at times that things in

---

[175] It might be noted here that in the discussions of powers in Kant's Leibnizian and Wolffian predecessors, the primary example of a power is a *vis repraesentativa*, which is mental power of representation in a monad. Thus Kant's idea of non-physical powers is very much in keeping with the idea of powers in his predecessors.

[176] See Descartes, *Meditations on First Philosophy*, in *The Philosphical Writings of Descartes* and *Correspondence*, vol. 2, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch (Cambridge: Cambridge University Press, 1984), Meditations II and VI.

[177] See also Gassendi's objections to Descartes' Sixth Meditation in *Objections and Replies*, in *The Philosphical Writings of Descartes* and *Correspondence*, vol. 2, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch (Cambridge: Cambridge University Press, 1984),

themselves ground appearances. He writes, for example, that it would be absurd to posit an appearance without something that grounds this appearance (B xxvii). And in the *Prolegomena*, he writes:

> [I]f we view the objects of the senses as mere appearances, as is fitting, then we thereby admit at the very same time that a thing in itself underlies them, although we are not acquainted with this thing as it may be constituted in itself, but only with its appearances, i.e., with the way in which our senses are affected by this unknown something. (AA IV: 314–5)

And to quote again a passage that has been discussed at length already, Kant writes: "The transcendental object that grounds both outer appearances and inner intuition is neither matter nor a thinking being in itself, but rather an unknown ground of those appearances that supply us with our empirical concepts of the former as well as the latter" (A 379f.). To understand the sense in which things in themselves, which are neither mental nor physical, ground mental and physical appearances, it might be illuminating to consider the similarity of Kant's view with a version of contemporary property dualism. Property dualism holds that mental and physical properties are heterogeneous and distinct but that these properties are grounded in something more fundamental. On Davidson's anomalous monism, for example, mental and physical properties both have a physical basis to which they can be reduced. Similarly, for Kant, mental and physical appearances are heterogeneous and are ontologically dependent on more fundamental things in themselves.[178] In contrast with Davidson's anomalous monism, however, Kant is offering a view according to which mental and physical appearances are grounded in fundamental things in themselves which are intrinsically neither mental nor physical.[179] Although it might be objected that this kind of neutral monism entails

---

[178] Otfried Hoeffe also argues that Kant is a monist. Although we agree Kant was a monist, we disagree about the relationship between this monism and property dualism. Hoeffe holds that Kant maintains "property monism," "as a methodological postulate," which denies that "mental and physical properties are essentially different in kind" (p. 269). Kant however maintains a "dualism of nature and freedom" (p. 270). See Otfried, Hoeffe, *Kant's Critique of Pure Reason: The Foundations of Modern Philosophy* (Dordrecht: Springer, 2010) pp. 266–270; originally published in German as *Kants Kritik der reinen Vernunft: Die Grundlegung der modernen Philosophie* (München: C. H. Beck, 2003). As I have shown, however, although mental and physical appearances are homogeneous in the sense that they are both appearances, Kant's point is that they are heterogeneous and grounded in things in themselves that are homogeneously neither mental nor physical.

[179] This is not to say that Kant is proposing a form of Spinozism according to which a single neutral substance, God, is the ground of mental and physical substances. Rather, there is a single kind of thing that grounds appearances. Kant criticizes Spinoza's "hyperphysical idealism" for claiming that the purposive unity of nature is grounded in a hyperphysical being

panpsychism or phenomenalism because mental and physical properties must have grounds of a like kind, Kant rejects such objections. In the Second Paralogism, for example, Kant explicitly rejects the claim that the unity of mind must be grounded in something that is itself simple and mental, arguing instead that the unity of thought may be an emergent property of something that is not itself capable of thought.[180] Analogously, physical and mental appearances may emerge from a ground that is in itself neither mental nor physical.

4.3.2 Interaction

Now that we have a sense of the ontological structure that Kant has in mind when he argues that for all we can possibly know interaction is possible because the ground of mental and physical appearances may be intrinsically neither mental nor physical, we may consider how Kant may have thought such interaction is supposed to take place.[181] And it must be admitted that Kant does not say a great deal explicitly about this issue. He appears to claim that the interaction of mental and physical appearances is to be explained in terms of the interaction of the neutral grounds of these appearances. But it is unclear in what sense the interaction of appearances is grounded in the interaction of things in themselves. The most obvious sense in which such an explanation would be plausible is if appearances are identical with their neutral grounds such that any change in the appearances would entail a change in their

---

(God) who creates the illusion of purpose in the *Critique of the Power of Judgment*; see AA 5:391–395. Galen Strawson also attributes a neutral monist view to Kant, although he does not provide a detailed interpretation of Kant's views on mind-body interaction. See *Mental Reality,* second edition (Cambridge: MIT Press, 2010), pp. 96–99. For a discussion of neutral monist positions, see: C.D. Broad, *The Mind and Its Place in Nature* (1925) (Oxon: Routledge, 2001), pp. 610–11 and pp. 632–640; Leopold Stubenberg, *Consciousness and Qualia* (Philadelphia & Amsterdam: John Benjamins Publishers, 1998); Bertrand Russell, *The Analysis of Matter* (London: Kegan Paul, 1927). Kant's neutral monism as I have interpreted it is similar to the current two-aspect theories discussed by Chalmers and others, which hold that there are neutral substances that can present themselves under the aspect of the mental or the physical; see David Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996).

[180] See A 351– A 354.

[181] Karl Ameriks has also argued that mental and physical appearances interact in virtue of the their homogeneous noumenal grounds. He justifies this in part on the basis of Kant's view in the lectures on metaphysics that interaction can occur only among like kinds. However, Ameriks construes these grounds only as non-material and so overlooks Kant's neutral monist claim that things in themselves are intrinsically neither mental nor physical. He also does not provide an account of how interaction is supposed to occur on Kant's view. See Karl Ameriks, *Kant's Theory of Mind*. *An Analysis of the Paralogisms of Pure Reason*, 2nd edition (New York: Oxford University Press, 2000), pp. 89–92.

neutral grounds and vice versa. But the fact that appearances are spatial and temporal and things in themselves are not means that appearances and things in themselves cannot be identical in this strict sense.[182] It might be argued instead that the type of relation Kant has in mind between appearances and their neutral grounds is a non-reductive supervenience relation. On such a supervenience relation, *A* supervenes upon *N* just in case there can be no change in *A* without an accompanying change in *N*. In this sense, *A* is ontologically dependent upon *N*. In Kantian vocabulary, this would mean that there can be no change among appearances without a corresponding change among things as they are in themselves.[183] Given this conception of supervenience, it would also appear that mental and physical appearances are epiphenomenal in the sense that the grounding relationship between appearances and things in themselves is not symmetrical. Although appearances may interact in virtue of their neutral grounds, they do not themselves cause any changes in their ground.[184] But the fact that the interaction of appearances is epiphenomenal does not imply that the interaction of appearances is not genuine. Appearances do genuinely causally interact, but they do so only in virtue of the interaction of their neutral grounds.

Although the appeal to supervenience and epiphenomenalism may help to describe the relation that obtains between mental and physical appearances and their noumenal grounds, it does not yet explain how mental and physical appearances could interact. And it appears that such an explanation would be needed in order to provide a detailed account of how mind-body interaction is possible. In a passage connected with his response to the dualist in the B edition of the *Critique of Pure Reason*, Kant provides some insight into how this interaction might be explained. He writes:

---

[182] This is not intended as an argument against a one-world view of transcendental idealism, which might hold that appearances and things in themselves are two aspects of one kind of thing. On the identity of appearances and things in themselves, see Nicholas Stang, "The Non-Identity of Appearances and Things in Themselves," *Noûs* 47(4) (2013), pp. 106–136.

[183] A pairing problem may arise for this account if appearances are not identical with things in themselves. Why is it that some particular appearance supervenes on some particular thing in itself rather than another? Kant cannot appeal as is customary to the fact that an appearance and its ground are spatially co-located, since appearances are spatially located whereas things as they are in themselves are not. On the pairing problem, see Jaegwon Kim, *Physicalism or Something Near Enough* (Princeton: Princeton University Press, 2005), ch.3, and Kim, *Philosophy of Mind*, 2nd edition (Cambridge: Cambridge University Press, 2006), pp. 44–50.

[184] Kant is quite clear about the asymmetry of this relation, and it is central, for example, to his account of freedom of the will. If appearances could interact with things in themselves, then our causally determined actions could causally determine our noumenally free actions, which would mean that such actions are in fact causally determined.

> But if one considers that the two kinds of objects [mental and physical substances] are different not inwardly but only insofar as one of them appears outwardly to the other, hence that what grounds the appearance of matter as a thing in itself might perhaps not be so different in kind, then this difficulty vanishes, and the only difficulty remaining is that concerning how a community of substances is possible at all, the resolution of which lies entirely outside the field of psychology, and, as the reader can easily judge from what was said in the Analytic about fundamental powers and faculties, this without any doubt also lies outside the field of all human cognition. (B 427–8)

The only difficulty that remains according to Kant after we recognize that the interaction of mental and physical appearances is grounded in the interaction of neutral things in themselves is to explain how substance interaction in general is possible. Although Kant argues that such an explanation lies outside the field of human cognition insofar as we cannot cognize things as they are in themselves, he nevertheless recognizes that such an explanation would require an appeal to fundamental powers. Despite his reservations about offering such an explanation here, his other comments on fundamental powers provide an insight into how such an explanation might be developed.[185] In his discussions of Wolff and Baumgarten on powers in his lectures on metaphysics, Kant endorses the idea that substances exhibit certain properties rather than others in virtue of their powers, although we cannot cognize these powers directly.[186] Similarly, Kant may argue that neutral things in themselves have powers whereby they bring about the exhibition of certain properties rather than others in appearances, although we can have no direct cognition of these powers. Thus an object has the property of being material and spatially located in virtue of the powers of the things in themselves that ground these properties. The existence of powers in things in themselves also explains how things in themselves interact with one another.[187] In contrast with Leibniz, who maintains that monads exercise their powers independently of one another in a universal harmony where their activities mutually reflect one another, Kant may suggest that the

---

[185] I attempt here only to explain how such an account might be developed on Kant's view not whether he believes the development of such an account would be legitimate given the restrictions he appears to place on our knowledge of things in themselves.

[186] On causal powers, see Kant's Metaphysik L_2 (Pölitz) (1790–1791?), AA 28:564.

[187] In a *Reflexion*, Kant also suggests that mind-body interaction is due to an "*urspruengliche Kraft*." See R 5457, AA 18:157f. See also Heinz Heimsoeth, *Studien zur Philosophie Immanuel Kants*: *Metaphysische Ursprünge und Ontologische Grundlagen* (Köln: Kölner Universitäts Verlag, 1956), p. 147.

interaction among things in themselves in virtue of their causal powers is reciprocal and therefore that things in themselves stand in a community of interaction.[188]

However, although things in themselves may be in reciprocal, causally dependent relations and ground the interaction of mental and physical appearances, the kind of causation Kant has in mind is not the same as that which applies to spatial and temporal appearances. Kant argues that the schematized category of causation cannot legitimately be applied to nonspatiotemporal things in themselves but that the unschematized category of ground and consequence may be applied to things in themselves.[189] As in the case of the non-temporal conception of mental properties, the notion of causation as a ground and consequence relation is conceived of as non-temporal. Moreover, because things in themselves are nonspatiotemporal, different laws than those that govern spatial and temporal appearances govern things in themselves. Kant writes, for example, that "[t]he lawfulness of things in themselves would necessarily pertain to them even without an understanding that cognizes them" (B 164), which suggests that things in themselves do exhibit some lawful behavior. And in the *Prolegomena*, he writes that the principles of connection of appearances in the sensible world are valid only for experience and not for things in themselves (AA 4:339–40). A thorough account of the laws or the "lawfulness" that governs or describes the interaction of things in themselves depends, however, upon a thorough account of things in themselves and their causal powers, an endeavor that Kant argues is beyond the scope of our epistemic capacities.

If this explanation of how appearances may interact in virtue of their grounds is correct, then we have a somewhat abstract characterization of the possibility of mind-body

---

[188] For a similar proposal, see Eric Watkins, "Kant's Model of Causality: Causal Powers, Laws, and Kant's Reply to Hume," *Journal of the History of Philosophy* 42(4) (2004), pp. 449–488. This is a position Kant also maintains in his account of mind-body interaction in the *New Elucidation* (AA 1:415), although it is clear that Kant distances himself in his critical period from his pre-critical claim that the reciprocal relation among substances depends on God (AA 1:415). On Kant's pre-critical views of mind-body interaction, see: Andrew N. Carpenter, "Kant's First Solution to the Mind/Body Problem," in *Kant und die Berliner Aufklärung*, vol. 2, ed. V. Gerhardt, R. Horstmann, & R. Schumacher (Berlin: De Gruyter, 2001), pp. 3–12; Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford University Press, 2000).

[189] Regarding non-temporal causation, Kant writes: "I speak of the mechanism of nature where the causality of the cause of an occurrence [*Begebenheit*] is itself an occurrence; and this is how it is with everything which happens insofar as the cause is an appearance; but insofar as the cause is a thing in itself, then the causality is not itself an occurrence, for it does not arise in time" (R 5978, AA 18:413). For related remarks see: R 5608, AA 18:249–51; R 5962, AA 18:401–5; A 544/ B 572; A 553/B 581; *Prolegomena*, AA 4:344, 346.

interaction that may be reconstructed from Kant's response to the dualist. Mind and body are appearances that are grounded in things in themselves which are intrinsically neither mental nor physical. And mental and physical appearances interact in virtue of the interaction of their neutral noumenal grounds. However, in order to make the interpretation more plausible, it may help to consider in more detail what is meant by mind-body interaction and how it is possible on this interpretation. There are a couple of ways to understand what the issue is concerning mind-body interaction or the "community in which our thinking subject stands to things outside us" (A 389) in Kant: (1) the first concerns how the mind can be causally affected by physical objects such that it is put into certain representational states; (2) the second concerns how mental causation is possible, i.e. how the mind can effect changes in the body associated with it through some kind willful action.

Regarding (1) how the mind may be affected by physical objects such that is put into some representational state, we might consider the following scenario. In this scenario, S_a and O_a refer to the subject and physical object as appearance respectively and S_n and O_n refer to the subject and object as noumenon or thing in itself respectively. S_a has a perception of a house occasioned by some empirical object O_a. The problem of interaction is that it is unclear how the house perception arises on the occasion of the subject being appropriately located with respect to the house since it is unclear how physical objects can cause mental representations. On the interpretation I am proposing, we should simply understand the scenario in the following way. S_n has a perception of a house caused by some noumenal object O_n. The causing of the perception of a house in this scenario is unproblematic because both S_n and O_n are not distinct in kind and can interact in virtue of their powers. S_a's state of perceiving the house supervenes on the states of S_n occasioned by O_n, which means there can be no change in S_a's perceptual states without a change in S_n's states.[190] This also means that S_a's states are epiphenomenal in the sense that they cannot have any effect on the state of S_n or O_n.[191] Regarding (2) the possibility of mental causation, we might consider the following

---

[190] S_n's states can change through affection by O_n or through its own noumenal freedom.

[191] One might wonder here about the so-called problem of double affection in Kant. The problem arises because of Kant's apparent commitment to both noumenal affection and empirical affection. If noumenal affection is true, then it seems that there is no role for empirical affection to play in causing representations. On the interpretation I have been developing, there is only noumenal affection, and empirical affection supervenes on noumenal affection, so there is no genuine problem of double affection. On the problem of double affection, see Nicholas Stang, "Who's Afraid of Double Affection," *Philosophers' Imprint* (forthcoming).

scenario. S_a has a mental state of desiring to pick up an object O_a. Since S_a's states are mental, they are different from O_a, so mental causation does not appear possible. However, on the view I am attributing to Kant S_n's mental states can determine the object O_n since S_n and its states and O_n are of the same kind and can interact in virtue of their powers. The change in states of S_a and O_a supervene on the change of states in O_n effected by S_n.

Now one problem with the interpretation above is that it does not appear to show why the neutral monist interpretation I have proposed has any virtues above and beyond the phenomenalist interpretation. In the characterization of (1), O causes some perceptual state in S. If one characterizes this as an interaction between O_a and S_a, then it is unclear why interaction is unproblematic since both O and S are construed as appearances, i.e. as mental items. The same is true of (2) where S causes some effect in O. If both O and S are appearances, then there is no reason to think that interaction is a problem. O_a and S_a are not heterogeneous but are the same kind of mental thing, an appearance. Moreover, the fact that Kant allows for the application of the category of cause to appearances but not things in themselves suggests that it makes more sense to say that O_a causes some state in S_a. But one problem with the phenomenalist interpretation is to understand what it means for S_a to be an appearance without being grounded in some thing in itself S_n. One sense in which S_a may be an appearance is that it appears to itself in some way as a mental item. But in order for this to be the case, as Kant says, the subject must somehow affect itself.[192] And since things in themselves affect appearances, then S_n must affect S_a. So there is more to the story than merely S_a and O_a even on the phenomenalist account. So the phenomenalist account would need to explain how S_n is affected by O_a such that it can effect a change in the state of S_a without appealing to some interaction between S_n and O_n upon which the change of states in S_a supervene. And if it can, then it must explain what the interaction of S_n and O_n consists in and whether S_n and O_n are of the same kind. But if they are of the same mental kind or physical kind, then this appears to violate Kant's claim that mental and physical properties do not attach to noumena. And if they are neither mental nor physical, then it appears that the proponent of the phenomenalist interpretation must accept the neutral monist interpretation.

The phenomenalist interpretation also appears to meet with some complications that arise in the epistemological context. One might wonder, for example, what makes S_a's

---

[192] On the problem of self-affection, see B 155–58. See also Corey W. Dyck, "Empirical Consciousness Explained: Self-Affection, (Self-)Consciousness and Perception in the B Deduction," *Kantian Review* 11(2006), pp. 29–54.

representation of O_a true or false? There appears to be no mind-independent fact against which S_a's representations can be measured. One might deal with this issue by appealing to a coherentist theory of truth or by suggesting that representations can in some regard be objective. But the neutral monist interpretation provides a much easier and intuitive answer. On the neutral monist interpretation, S_a's representation of O_a are true or false because S_a's representations supervene on S_n's representation of O_n. Of course, this leaves open the possibility that S_n may falsely represent O_n in which case S_a's representation of O_a would be false. A thorough account of the epistemological issues is, however, beyond the scope of this chapter. Having shown that the neutral monist interpretation of Kant's proposed solution to the problem of the possibility of mind-body interaction has some plausibility, we may now turn to some objections that might be raised against the interpretation and Kant's account itself as I have interpreted it.

4.4 Objections and Replies

One objection that may be raised against Kant's account as I have interpreted it is that in maintaining that interaction is unproblematic because it is possible that heterogeneous mental and physical appearances interact in virtue of their neutral grounds, the view overlooks some additional hindrances to interaction. Although it may be granted that Kant maintains that things in themselves are neutral in the sense that they are intrinsically neither mental nor physical, and so are homogeneous in this regard, they may nevertheless be heterogeneous in other regards. Things in themselves may conceivably have heterogeneous properties such as being quizzical or topsy-turvy, and the heterogeneity of these properties may make them incapable of interaction, and thus also make the interaction of mental and physical appearances impossible. However, it might be responded that in providing an account of the possibility of mind-body interaction, Kant is merely concerned with showing that the heterogeneity of mental and physical appearances is no hindrance to their interaction. Whether there may be other reasons to believe that things as they are in themselves are not capable of interaction is something that Kant does not appear concerned with in his discussion of the problem of mind-body interaction. He appears to believe that things in themselves are homogeneous in all respects that are relative to their interaction and that the interaction of things in themselves is therefore unproblematic.

A second objection is that this interpretation of Kant's account is unable to account for Kant's proposed solution to the problem of freedom of the will. His solution holds that as appearances we are subject to thoroughgoing determinism, but as things in themselves we are free. This is because the category of causation involved in causal determinism does not apply to things in themselves but only to appearances. As things in themselves, however, it is conceivable that we are capable of acting freely and in accord with our choice of an intelligible character from which our empirical actions flow. However, this conception of a noumenal self as freely acting requires that it have some form of mentality. This is to say, it would make little sense to say that persons as things in themselves are free and thus morally responsible for their choice of intelligible character if they do not have the capacity to choose this character. Since this is the case, we should hold that Kant requires a ground of appearances that is itself mental. The above account is, however, able to accommodate this insight. As was argued, Kant maintains only that things in themselves do not have spatial and temporal mental and physical properties and that mental and physical properties do not apply univocally to appearances and things in themselves. So although it may be conceded that things in themselves do not have the capacity to deliberate about actions in time, it may be argued that certain things in themselves nevertheless have the capacity for making a-temporal or timeless choices. Although this may seem far-fetched, this is exactly what Kant's account of freedom appears to require. Kant writes, regarding the noumenal self, for example: "this acting subject, in its intelligible character, would not stand under any conditions of time, for time is only the condition of appearances but not of things in themselves" (A 539/ B 567).[193] For if the choice of character occurred in time, it would also be subject to causal determinism, which is contrary to Kant's proposed solution to the problem of freedom of the will. I do not wish to argue that this is in fact Kant's account of freedom but only that the account of Kant's proposal regarding the possibility of mind-body interaction is at least compatible with one of the major interpretations of Kant's account of freedom that has been proposed.

A third objection that might be raised is that the view I am attributing to Kant violates his claim that we cannot have knowledge of things as they are in themselves. In the *Critique of Pure Reason*, Kant claims that we cannot have non-analytic knowledge of things as they are in themselves. But the view I have attributed to Kant suggests that Kant maintains that

---

[193] See Allen Wood's interpretation of Kant on freedom of the will in "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 73–101.

things in themselves are neither spatial nor temporal. In response it might be argued that the claim that things in themselves are neither spatial nor temporal follows analytically from things that Kant maintains that we do know. He argues explicitly that only appearances can be spatial and temporal. So it follows that if things in themselves are not appearances, they are not spatial or temporal. I have also pointed out that Kant maintains that things in themselves affect us, i.e. that changes in appearances supervene on changes in things in themselves. And as I have shown, Kant does say this. But this latter point seems to violate the strictures on knowledge of things in themselves, particularly since interaction does not follow analytically from anything Kant claims about appearances. At this point, one might simply suggest that Kant is inconsistent about the limitations on our knowledge of things in themselves and the kinds of claims he does make about things in themselves. However, it should also be pointed out that the interpretation I have been providing does not claim explicitly that Kant maintains that we can know that things in themselves are non-spatial and non-temporal, nor that the states of appearances supervene on the states of things in themselves, nor that things in themselves interact in virtue of their powers. As such, the interpretation does not violate Kant's strictures on our knowledge of things in themselves. The claim has simply been that if transcendental idealism and Kant's claims about things in themselves are interpreted in the way I have proposed, then Kant has provided a positive response to the dualist regarding how mind-body interaction is possible. Whether the account of transcendental idealism I have attributed to Kant is correct and how it accords with his restrictions on knowledge of things in themselves is a separate question.

Having cleared away some of the hindrances to the neutral monist interpretation of Kant's account of interaction, we may consider some of its virtues. Interestingly, Kant's account of the possibility of mind-body interaction as I have interpreted it has the potential to avoid some typical objections to property dualism. Physicalist property dualism, which maintains that there are irreducible mental and physical properties that are grounded in a more fundamental physical basis, suffers from the problem that mental properties do not appear to be causally efficacious. Since the physical world is causally closed, the mental can have no effect on its physical basis. Kant's account as I have interpreted it is not susceptible to such a worry, however, because mental and physical properties are grounded in neutral things that are capable of interaction. So mental and physical appearances can mutually interact through the interaction of their neutral grounds. Likewise, Kant's account as I have interpreted it also escapes the problem of the overdetermination of physical events by physical and mental causes that was well known to Leibniz and others and has more recently

been raised for property dualism.[194] Since mental properties are active only in virtue of their physical grounds, any event of mind-body interaction appears overdetermined in the sense that it has both a physical cause and an unnecessary mental cause. Again, Kant's account as I have interpreted it does not fall prey to this objection because the mental and physical can interact only in virtue of the interaction of their neutral grounds. Although such solutions come at the cost of accepting transcendental idealism, Kant's account nevertheless represents an interesting contribution to the historical debate about mind-body interaction.

4.5 Conclusion

In the foregoing discussion, we have seen that Kant argues that mind-body interaction is problematic for the substance dualist because the dualist takes mental and physical substances to be things as they are in themselves rather than appearances. One line of interpretation claims that Kant argues that we are ignorant of things as they are in themselves and so are therefore ignorant of whether and how mental and physical substances as they are in themselves may be capable of interaction. This anodyne response is, however, unsatisfying because it fails to take seriously Kant's appeals to mental causation in his account of freedom of the will and his account of mental affection by things in themselves, both of which require an explanation of the possibility of mind-body interaction. I have argued for an interpretation of Kant's critique of the dualist and proposed solution to the problem of the possibility of mind-body interaction that overcomes the problems with the anodyne response. On the interpretation I have proposed, Kant argues that mental and physical appearances are grounded in things in themselves, which are neither mental nor physical, and that these appearances interact in virtue of the interaction of the things in themselves that ground them. In the next chapter, we will consider Kant's account of the compatibility of freedom of the will and determinism.

---

[194] See the discussion of overdetermination and Kant's predecessors in Desmond Hogan, "Noumenal Affection," *Philosophical Review* 118(4) (2009), p. 511.

Chapter 5

Kant's Compatibilist Theory of Freedom of the Will

5.1 Introduction

In the previous chapters, we have seen that Kant mounts a criticism of rationalist claims regarding our knowledge of the soul and its fundamental powers. He shows that thought is grounded in a thing in itself and that it is possible that the ground of thought possesses multiple powers. We have also seen that Kant argues for a positive view of the nature of personhood and the possibility of mind-body interaction. In this chapter, we will consider another important aspect of Kant's metaphysics of mind and his critique of rational psychology, which also plays a role in his account of moral responsibility. In particular, we will consider the role of our capacity for reason in Kant's attempt to reconcile freedom of the will and causal determinism.

In the Third Antinomy of the *Critique of Pure Reason*, Kant argues that the apparent conflict between incompatibilist determinism, which holds that all events are determined by antecedent events, and incompatibilist libertarianism, which holds that an action can be initiated independent of antecedent events, arises because each position takes appearances to be things in themselves. According to Kant's "critical solution," which "does not consider the question objectively at all, but instead asks about the foundations of the cognition in which it is grounded" (A 484/ B 512), if we accept transcendental idealism and its distinction between appearances and things in themselves, then it can be shown that "freedom and natural necessity in one and the same action" do not "contradict each other" (A 557/ B 585) and are therefore compatible. Although commentators agree on the broad outlines of Kant's proposed solution, they have often disagreed about whether Kant is providing a metaphysical account of the compatibility of freedom and determinism or some other kind of account.

According to Allen Wood's representative metaphysical interpretation, freedom and determinism are compatible because "the self as free moral agent belongs to a different world

from that of the self as natural object."[195] On this interpretation, we have an empirical self that is an appearance and exists in space and time, and we have an intelligible self that is a thing in itself that exists outside of space. Since the intelligible self exists outside of space and time, its actions are free from causal determinism, and since the empirical self exists within space and time, its actions are subject to causal determinism. When a person acts, they freely choose an intelligible character, which then wholly determines their empirical character and the empirical actions that flow from this character (A 539/B 567). This choice of an intelligible character is free from causal determinism since it is a choice made by an intelligible self. A person is also considered free because for any given action, they could have done otherwise. They could have done otherwise empirically if and only if they had chosen a different intelligible character. And the choice of a different intelligible character from a set of possible alternatives would have lead to a different empirical character and hence also to different course of empirical actions. In order to explain how our intelligible character remains causally efficacious in empirical actions, Wood also maintains that the choice of intelligible character is "considered simultaneous with each act as it occurs in the temporal order," or in other words, that our intelligible character is capable of "timeless agency."[196] If our agency were not timeless in this way, it would be subject to causal determinism. Building upon this metaphysical interpretation, some commentators have also argued that the ability to choose one's intelligible character, and therefore also the ability to have done otherwise, also requires the ability to change the laws of nature or the ability to create a miracle as an exception to the existing laws of nature. Since the laws of nature determine the course of empirical events, it is thought that any intervention in the course of these events through one's timeless intelligible agency requires that a miracle be produced as an exception to the laws of nature or that the laws of nature themselves must be altered through one's intelligible choice. Since Kant is unlikely to countenance miracles, it is argued that we must have the ability to change the laws of nature governing appearances through our choice of intelligible character. Both the claim that we have timeless agency and that we have the ability to change the laws of nature have been met with skepticism. One reason for this skepticism is that it appears incoherent to say that a timeless act is simultaneous with anything since only something in time can be simultaneous with something else. And as

---

[195] Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 73–101.
[196] Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), p. 96.

many have remarked, the view also appears to entail that one may be held morally responsible for all of the actions that flow from one's choice of intelligible character and the accompanying alteration of natural laws including actions that follow from one's choice but also those that preceded it. The problems related to timeless agency, our ability to alter the natural laws, and the extent of our moral responsibility have led Jonathan Bennett to quip that "most of us have long thought it [Kant's theory of noumenal freedom] to be dead, and after […] restorative measures the corpse still refuses to stir."[197]

Because of these problems Henry Allison, Hud Hudson and others have attempted to restore the corpse of Kant's account of freedom by adopting a methodological two-aspect account of Kant's distinction between appearances and things in themselves that does not hold that there is a causal relation between one's intelligible character and one's empirical acts that is based in noumenal freedom.[198] Hudson argues, for example, for an anomalous monist interpretation of Kant's proposal.[199] On Hudson's interpretation, empirical actions are subject to thoroughgoing causal determinism because they are subject to natural laws, and intelligible actions are not subject to thoroughgoing causal determinism, and are therefore free, because they are not subject to natural laws. This leads him to claim that when we are described as intelligible beings not governed by natural laws we are considered free, and when we are described as empirical beings according to natural laws we are subject to causal determinism. Although such an interpretation allows one to avoid the unsavory metaphysical issues regarding timeless agency and the ability to change the natural laws, there are nevertheless problems with the interpretation. First, if one thinks that Kant's disagreement

---

[197] Jonathan Bennett, "Kant's Theory of Freedom," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), p. 107.

[198] Henry Allison has also argued for what might be construed as a two-aspect interpretation of Kant's solution to the problem of freedom of the will, although he argues that Kant is an incompatibilist. See Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp. 1–2, 29–46.

[199] See Hud Hudson, *Kant's Compatibilism* (Ithaca: Cornell University Press, 1994); see also "Kant's Third Antinomy and Anomalous Monism," in *Immanuel Kant: Groundwork of the Metaphysics of Morals in Focus*, ed. Lawrence Pasternack (London: Routledge, 2002), pp. 234–267. For a similar approach see Ralf Meerbote, "Kant on the Nondeterminate Character of Human Actions," in *Kant on Causality, Freedom, and Objectivity*, ed. W. L Harper and Ralf Meerbote (Minneapolis: University of Minnesota Press, 1984), pp. 138–63, and "Kant on Freedom and the Rational and Morally Good Will," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 57–72. Donald Davidson's anomalous monism is in part inspired by his view that for Kant "freedom entails anomaly." See "Mental Events" (1970), in *Essays on Actions and Events*, 2nd ed. (Oxford: Clarendon Press, 2001), p. 206.

with the incompatibilist views of his early-modern predecessors is not merely a verbal dispute about how best to understand the words "freedom of the will" and that it truly concerns the question of whether freedom of the will is *in fact* possible and compatible with determinism, then it is simply inadequate for Kant to argue that we are free when regarded in one way and causally determined when regarded in another. If one wants to know whether you have enough money in your pocket to cover your bar tab, then it would do you no good to say that if you are considered with twenty dollars in your pocket you have enough, but considered without twenty dollars in your pocket you do not. The terms in which the situation is described have no bearing on what is in fact the case. Second, Hudson's anomalous monist interpretation holds that the ontological ground of mental and physical properties is in fact physical and that mental events are token-token identical with physical events. But if this is true, then it simply seems to be an error to ascribe freedom to mental events because such events are token-token identical with physical events that are causally determined according to natural physical laws. Kant's account of freedom of the will also requires that intelligible causes are themselves uncaused, but as numerous commentators have pointed out, an event that is caused under one description and uncaused under another is not uncaused.[200] Such problems suggest that the methodological interpretations have not fared much better than the metaphysical interpretations in their necromantic endeavors.

In what follows, I will revisit Kant's apparent commitment to timeless agency and the idea that our ability to do otherwise entails the ability to change the laws of nature in order to provide support for a metaphysical interpretation of Kant's account of freedom of the will and determinism. In section 5.2, I argue that Kant maintains that our freedom of the will is grounded in the intelligible capacity to act from reasons and that this capacity produces effects in the empirical world. Although such a capacity may be masked or undermined by determinate causal events, Kant maintains that we have freedom of the will so long as we retain the intelligible capacity to act from reasons, which is grounded in our power of spontaneity. And I show how this conception allows Kant coherently to maintain that our capacity for reason is in some regard timeless. In section 5.3, I consider attempts that have

---

[200] See Tobias Rosefeldt, "Kants Kompatibilismus," in *Sind wir Bürger zweier Welten?: Freiheit und moralische Verantwortung im transzendentalen Idealismus*, ed. M. Brandhorst, A. Hahmann, B. Ludwig (Hamburg: Felix Meiner Verlag, 2012), pp. 77–109, p. 6. See also Wolfgang Ertl, "Hud Hudson: *Kant's Compatibilism*," *Kant-Studien* 90 (1999), pp. 371–384, and Derk Pereboom, "Kant on Transcendental Freedom," *Philosophy and Phenomenological Research* 73 (2006), pp. 537–567, footnote 4.

been made to make sense of Kant's commitment to our ability to change the laws of nature. And I argue that Kant's account as I have interpreted it does not require such an ability in order for freedom and natural necessity to be compatible. In section 5.4, I show the implications of this interpretation for questions regarding the kinds of entities that have freedom of the will and the extent of our moral responsibility. And section 5.5 concludes by summing up the main points of the chapter.

5.2 Timeless Agency and The Capacity for Reason

In his account of freedom of the will, Kant accords an important place to our spontaneity and the spontaneous faculty of reason. For Kant, our cognition arises from two basic sources of the mind. He writes:

> Our cognition arises from two fundamental sources in the mind [Grundquellen des Gemüts], the first of which is the reception of representations (receptivity of impressions), the second the faculty for cognizing an object by means of these representations (spontaneity of concepts); through the former the object is given to us, through the latter it is thought in relation to that representation (as a mere determination of the mind). […] If we will call the receptivity of our mind to receive representations insofar as it is affected in some way sensibility, then on the contrary the faculty for bringing forth representations itself, or the spontaneity of cognition, is the understanding. (A 50f./ B 74f.).

Kant argues in the Transcendental Deduction that we have certain capacities or faculties (*Vermögen*) that have their source in our ability to be affected and to bring forth representations. He often equates our ability to be affected with the capacity for sensibility and the ability to generate representations with the capacity for understanding. Receptivity and spontaneity and the various mental capacities they afford work together to allow us to perceive objects and to organize our experience through judgments involving the categories.

When Kant mentions spontaneity and receptivity, he implicitly situates his discussion within a broader historical context in which spontaneity was regarded as a fundamental power of the mind or soul. Kant's discussion of spontaneity in his lectures on metaphysics and his attempt in the subjective deduction to distance himself from the rationalist claim that there is only a single fundamental power of the soul suggest that Kant was also well aware of this. In his discussion of spontaneity in Metaphysik L_1, he refers to his predecessors doctrines of spontaneity using the terminology of transcendental idealism. According to Kant "the soul is a being which acts spontaneously, simply speaking <*simpliciter spontan*>" and he goes on to suggest that this means that "the human soul is free in the transcendental sense <*in sensu*

transcendentali>" (AA 28:267). The transcendental concept of freedom means "absolute spontaneity, and is self-activity from an inner principle according to the power of free choice" (AA 28:267). In contrast with Wolff and his followers, Kant also argues that receptivity and spontaneity are independent powers of the mind that cannot be reduced to a single fundamental power of representation.[201] Kant also classifies our mental capacities according to whether they are spontaneous or receptive. Sensibility is characterized in terms of the power of receptivity, and the understanding is characterized in terms of the power of spontaneity. More importantly, however, although Kant mentions only that understanding is a spontaneous capacity in the Transcendental Deduction, it is also clear that our capacity for reason is also characterized in terms of the power of spontaneity. The capacity for reason arises from the power of spontaneity, which allows reason to produce representations without being determined through affection. The capacity for reason allows us to reason from premises to conclusions and to organize our thoughts in hierarchical structures. But it also allows us to provide moral maxims, laws, and imperatives with which reason demands we act in accordance. According to Kant, when we exercise our capacity for reason in order to reason from premises to conclusions or in order to give ourselves a moral imperative, we also recognize that this capacity for reason allows us to think in a way that is not antecedently determined by sensibility or the vicissitudes of desire or impulse.[202]

Importantly for Kant, we also recognize that in using our capacity for reason we are acting from our power of spontaneity. And the fact that we can act from spontaneity also shows according to Kant that we have freedom of the will in the sense that we possess the ability to begin a causal series independent of antecedent conditions. In the initial characterization of freedom in the Third Antinomy, Kant refers to it as an "absolute spontaneity of an action" (A 448/ B 476). Later it becomes clear, however, that it is reason itself that exhibits the kind of spontaneity associated with freedom. He writes, for example:

> [R]eason does not give in to those grounds which are empirically given, and it does not follow the order of things as they are presented in intuition, but with complete spontaneity it makes its own order according to ideas, to which it fits the empirical conditions and according to which it even declares actions to be necessary that yet

---

[201] Kant does speculate about whether such a fundamental unknown root of our mental powers exists, but the subjective deduction appears to show that we have several irreducible powers. See Corey W. Dyck, "The Subjective Deduction and the Search for a Fundamental Force," *Kant-Studien* 99(2) (2008), pp. 152–179.

[202] Reason is therefore an intelligible capacity rather than a merely sensible capacity. Kant discusses the spontaneity of reason and the understanding and the receptivity of sensibility at A 546f./ B574f.

> have not occurred and perhaps will not occur, nevertheless presupposing of all such actions that reason could have causality in relation to them; for without that, it would not expect its ideas to have effects in experience. (A 548/ B 576).[203]

Kant's understanding of reason as exhibiting spontaneity and thus the freedom to act independent of causal necessity is also not an odd formulation given Kant's discussion of spontaneity in the lectures on metaphysics where he also suggests that the soul exhibits freedom when it acts from the power of spontaneity.[204] In the Third Antinomy discussion, Kant has just provided more information about how we exhibit spontaneity through our capacity for reason.[205]

Kant illustrates the role of reason in our free actions using the example of a malicious lie. When considering whether a person is morally responsible for some action, one might consider their empirical character, i.e. their particular moral education, the company they keep, and their natural temperament, and attribute their immoral act to the fact that they were causally determined to act in the way they did because of their empirical character. Nevertheless, despite their empirical character and the various circumstances that may have determined them to act in a certain way, we hold them morally responsible for their actions. This practice of holding persons morally responsible "is grounded on the law of reason, which regards reason as a cause that, regardless of all the empirical conditions just named, could have and ought to have determined the conduct of the person to be other than it is" (A 555/ B 583). Regarding this judgment of imputation, Kant writes:

---

[203] See also the Groundwork, AA 4:448, where Kant suggests that reason must see itself as the author of its own principles.

[204] In his lectures on metaphysics, Kant distinguishes between *spontaneitas absoluta vel simplicter talis* and *spontaneitas secundum quid* (AA 28:267). In the former, spontaneity, which corresponds to the transcendental concept of freedom, is absolute, and in the latter, it is qualified in some respect, i.e. it is freedom under a condition. The latter is the "freedom of the turnspit," which he accuses Leibniz of proposing; see the *Critique of Practical Reason* (AA 5:97). See also the discussion in Stefanie Grüne, "Kant and the Spontaneity of the Understanding," in *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, ed. Dina Emundts (Berlin: Walter de Gruyter, 2013), p.149f.n.10.

[205] The literature on spontaneity in Kant is immense. One influential text is Robert Pippin, "Kant on the Spontaneity of the Mind," in *Idealism as Modernism: Hegelian Variations* (Cambridge: Cambridge University Press, 1997), pp. 29–55. Pippin also recognizes that for Kant spontaneity cannot be a feature of a phenomenal subject since all phenomena are subject to causal necessity (39). I would suggest therefore that spontaneity must be a property of a thing in itself. See also Henry Allison, "Autonomy and Spontaneity in Kant's Conception of the Self," in *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge, Cambridge University Press, 1996), pp. 129–142.

> In this judgment of imputation, it is easy to see that one has the thoughts that reason is not affected at all by sensibility, that it does not alter (even if its appearances, namely the way in which it exhibits its effects, do alter), that in it no state precedes that determines the following one, and hence that reason does not belong at all in the series of sensible conditions which make appearances necessary in accordance with natural laws. It, reason, is present to all the actions of human beings in all conditions of time, and is one and the same, but it is not itself in time, and never enters into any new state in which it previously was not. (A 555f./ B 583f.)

According to Kant, reason is free of determination by antecedent sensible conditions. Reason does not belong to the series of sensible conditions because it determines itself in accordance with logical principles and laws rather than in accordance with antecedent events and natural laws. In this regard, it stands outside of the determinate causal series. And since it stands outside of this determinate causal series, it is also free. Kant also argues that our capacity for reason, which is grounded in our power of spontaneity, is an intelligible cause of our actions. The idea appears to be that reason is the intelligible cause of our actions insofar as it determines an intelligible character from which our empirical character and our empirical actions flow.[206] To use the example of the malicious liar, the liar has a capacity to give herself a moral maxim, which holds for example that lying is permissible, which in turn determines her intelligible character and also the actions that flow from this intelligible character. Because we have the capacity to act from reason, we are not antecedently determined to act in a particular way by empirical events and so act freely. The liar freely determines her intelligible character and the actions that flow from this through her capacity for reason and her capacity to provide a moral maxim for herself. And we recognize, as Kant says, that all persons who possess this capacity for reason are able to give themselves moral maxims and determine their intelligible character and the actions that flow from this character and thus are free and so may be held morally responsible for their actions.

Some interpreters have argued that Kant's account of how we act freely using the capacity for reason is merely meant to show that we are free in a practical sense rather than free in a transcendental sense. Kant often speaks as though our capacity for reason and our capacity to establish moral maxims independent of antecedent sensible conditions show only

---

[206] Kant writes: "Pure reason, as a merely intelligible faculty, is not subject to the form of time, and hence not subject to the conditions of the temporal sequence. The causality of reason in the intelligible character does not arise or start working at a certain time in producing an effect. For then it would itself be subject to the natural law of appearances, to the extent that this law determines causal series in time, and its causality would then be nature and not freedom" (A 551f./ B 579f.).

that we have some practical warrant for regarding persons as free. We regard persons as free because they have the capacity for reason, but we have no reason to think that they are transcendentally free, i.e. free to bring about a causal series on their own independent of antecedently determining conditions, nor do we have any understanding of how such freedom might operate.[207] There is some evidence for such an interpretation in the Doctrine of Method where Kant holds that practical freedom requires the ability of reason to determine our will but does not require the independence of reason itself "from all determining causes of the world of senses" (A 803/ B 831), which is only a matter of theoretical concern. But in the discussion surrounding the Third Antinomy the stakes in Kant's discussion of reason and freedom of the will appear to be much higher. Here Kant goes so far as to argue that the fact that we have a capacity for reason even provides epistemic warrant for thinking that we are transcendentally free.[208] Whether Kant succeeds in demonstrating our freedom through this line of argument in the Third Antinomy is dubious.[209] And Kant himself derides the rationalists for a similar line of reasoning by which they attempt to establish the simplicity of the soul through empirical means. But whether or not Kant is in a position to show that we can know that we are transcendentally free on the basis of the fact that we exercise some degree of freedom in our capacity for reason, he nevertheless appears to maintain that we actually exercise our transcendental freedom through our capacity for reason insofar as reason is an intelligible cause that constitutes our intelligible character and operates independent of any antecedently determining conditions.[210]

---

[207] On Kant's practical and speculative demonstrations of our freedom, see Karl Ameriks, *Kant's Theory of Mind*, 2nd ed. (New York: Oxford University Press, 2000), pp. 193–196. As Ameriks sees it, Kant did not mean that practical freedom is sufficient for morality, but rather that we cannot have a proof of how absolute freedom works. Since we cannot have such a transcendental proof, we have to settle with one that is merely practically sufficient.

[208] In this regard, Kant comes very close to Tetens who also argues that we have an awareness of our freedom and spontaneity. See Johann Nicolas Tetens, *Philosophische Versuche über die menschliche Natur und ihre Entwicklung*, vol. 2 (Leipzig: 1777), p. 4. See also Bernd Ludwig's discussion of the development of Kant's arguments for freedom of the will in Bernd Ludwig, "Die '*consequente Denkungsart* der speculativen Kritik.' Kants radikale Umgestaltung seiner Freiheitslehre im Jahre 1786 und die Folgen für die Kritische Philosophie als Ganze," *Deutsche Zeitschrift für Philosophie* 58(4) (2010), pp. 595–628.

[209] For a discussion of Kant's argument, see Karl Ameriks, "Kant's Deduction of Freedom and Morality," in *Interpreting Kant's Critiques* (Oxford: Oxford University Press, 2003), pp. 161–192; and Ameriks "Kant's *Groundwork* III Argument Reconsidered, " in *Interpreting Kant's Critiques* (Oxford: Oxford University Press, 2003), pp. 226–248.

[210] It was not uncommon in the period to talk about moral character in dispositional terms. Eberhard, for example, in *Allgemeine Theorie des Denkens und Empfindens* (Berlin: 1776)

One problematic point in Kant's metaphysical account of how we act freely on the basis of our capacity for reason is that he appears to argue that reason in some sense allows us to choose the intelligible character from which our actions flow. In the case of the malicious liar, she chose certain maxims according to which to act and thus chose her intelligible character, which subsequently determined her empirical actions. She was also free to choose differently. As Kant himself recognizes, however, choices take place in time and so are subject to causal determinism, which would mean that the choice of an intelligible character is also causally determined and therefore that we lack freedom of the will. This problem has led commentators to argue that our choice of intelligible character must be timeless in the sense that we exercise a choice of intelligible character without this choice taking place in time in the way that empirical choices take place in time. And there have been a number of attempts to make sense of what this notion of timeless agency means and whether it is coherent.[211] There is certainly evidence in Kant's discussion that suggests he thought of our

---

refers to the "moral disposition (character) of a human being." See Eric Watkins (ed. and trans.), *Kant's Critique of Pure Reason: Background Source Materials* (Cambridge: Cambridge University Press, 2009), p. 323.

[211] Commentators have defended the notion of timeless agency and timeless choice in various ways. Wolfgang Ertl argues that Kant's compatibilism is coherent if one recognizes that he was proposing a view similar to the "Ewigkeitslösung" found in Molina and Boethius. See Ertl: "Schöpfung und Freiheit: Ein kosmologischer Schlüssel zu Kants Kompatibilismus," in *Kants Metaphysik und Religionsphilosophie* (Hamburg: Felix Meiner, 2004), pp. 43–76; *Kants Auflösung der "dritten Antinomie": Zur Bedeutung des Schöpfungskonzepts für die Freiheitslehre* (Freiburg, München: Karl Alber Verlag, 1998); "'Ludewig' Molina and Kant's Libertarian Compatibilism," in *A Companion to Luis de Molina*, ed. A. Aichele and M. Kaufmann (Cologne: Brill, 2013). Allen Wood argues that the choice of an intelligible character is timeless in the sense that it is simultaneous with one's empirical actions. As Hud Hudson rightly points out, this proposal is incoherent because if a choice is outside of time it cannot be simultaneous with anything. See Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 73–101, and Hud Hudson, *Kant's Compatibilism* (Ithaca: Cornell University Press, 1994), p. 26. Tobias Rosefeldt has argued that Kant's conception of agency is borrowed from Baumgarten and other rationalists according to whom action (*actio*) is the capacity for a substance to bring about a change in its accidents through its own power. In contrast with the idea of a change of states, the idea of actualizing an accident in a substance is not a temporal notion since Kant believes that substances can have properties without having them at a particular point in time. See: Tobias Rosefeldt, "Kants Kompatibilismus," in *Sind wir Bürger zweier Welten?: Freiheit und moralische Verantwortung im transzendentalen Idealismus*, ed. M. Brandhorst, A. Hahmann, B. Ludwig (Hamburg: Felix Meiner Verlag, 2012), pp. 88–89; and Alexander Baumgarten, *Metaphysica* (Frankfurt: 1757), §197, §210; Immanuel Kant, Metaphysik L_2 (AA 28:565). Rosefeldt's treatment also accords with my proposal that we manifest an intelligible character rather than choosing the character from among a set of possible intelligible characters. Kant also appears in the *Prolegomena* to offer such a conception of

choice of intelligible character as a timeless one that does not stand within the temporal order.[212] But it appears that the number of problems raised in trying to understand what a timeless choice or timeless action might be is a good reason to seek some other interpretation.

If one accepts the preceding interpretation according to which reason is a capacity that allows one to act according to certain moral maxims, then there may be a weaker way to understand Kant's apparent claim that we exercise timeless choice or timeless agency. I suggest that Kant's point regarding timelessness is only that our intelligible capacity for reason is timeless in the sense that any capacity is timeless. Consider a glass that has the capacity to break when it is thrown against the wall. The glass retains this capacity even when the capacity is not manifested. And the capacity can be said to be simultaneous with the existence of the glass. In a similar way, we retain the capacity to act according to reason in a non-determined way even when we do not manifest this capacity or even when this capacity is manifested in some deficient way. And this capacity is simultaneous with our existence. It is because Kant conceives of reason as a capacity that he suggests that "reason is thus the persisting condition of all voluntary actions under which the human being appears," and "reason is present to all the actions of human beings in all conditions of time, and is one and the same, but it is not itself in time" (A 556/ B 584). Reason is a capacity of persons, which they retain as a potential that can be manifested in various circumstances.[213] And it is timeless and simultaneous with the existence of a person in the sense that this capacity is retained during the course of the existence of a person regardless of whether it is manifested or not. This capacity for reason becomes subject to causal determinism when it is manifested in an empirical action. But as long as it remains merely a capacity, it is not subject to causal determinism. To use the example of the glass again, it has a timeless capacity to break, but this capacity only enters the series of determinate causal events when it is actually manifested and the glass breaks. Likewise, the liar has a capacity to give herself a moral maxim to lie or not lie, and this capacity enters the series of determinate causal events only when it is exercised in some way or another in an empirical action.

---

action when he writes that "the relation of an action to the objective grounds of reason is not a temporal relation" (AA 4:346).

[212] See for example A 552ff./ B 580ff.

[213] This account of the timelessness of our capacity for reason also explains why Kant says that we do not ask why the choice of intelligible character was not different, because there is no such choice, but we ask only why this capacity manifested itself in empirical nature in the way that it did. And Kant is quite explicit that "no answer to this is possible" (A 556/ B 548).

It is also crucial to Kant's account, as some have interpreted it, that our free actions from reason somehow determine the occurrence of empirical events through the determination of our empirical character. This is one way in which it is thought that Kant can show that intelligible causation from freedom and causation according to natural necessity are compatible. But I would like to suggest that Kant may be interpreted as making a much weaker point. He does not require that an intelligible cause must intervene to determine the course of events but only that the maxims provided by reason should be in accord with what is possible in the natural world. This is simply another way of formulating the dictum that "ought implies can." Kant writes for example that "now of course the action must be possible under natural conditions if the ought is directed to it; but these natural conditions do not concern the determination of the power of choice itself, but only its effect and result in appearance" (A 547f./ B 575f.) This is to say that our moral imperatives should be directed toward things that can be accomplished given the natural laws. And our judgments of moral responsibility should be in accord with this understanding of how our moral imperatives fit with empirical events and the laws of nature. For example, I should not give myself a moral maxim that maintains that anytime a bullet is fired at another person, I should catch the bullet in mid air in order to save them. It is impossible for me to implement this maxim given the laws of nature. And we should not hold someone morally responsible for the fact that they failed to choose to stop a bullet in mid air to save someone else because such a demand does not accord with the possibilities of nature.

However, despite Kant's warning that the demands of reason should be in accordance with the possibilities in nature, he nevertheless recognizes that reason does give imperatives for action even regarding situations that "have not occurred and perhaps will not occur" (A 548/ B 576).[214] There is a tendency for us to hold someone morally responsible for cases in which it may have been impossible given the laws of nature and the natural course of events for them to do otherwise. We demand that someone should not have lied in a certain situation, but it may have been physically impossible for her not to lie because each time she attempts to tell the truth her particular brain structure determines her to say the opposite of what she intend to say. Or she was simply unable to overcome her flawed upbringing or

[214] Kant also recognizes that sometimes the demands of reason do accord with nature: "At times, however, we find, or at least believe we have found, that the ideas of reason have actually proved their causality in regard to the actions of human beings as appearances, and that therefore these actions have occurred not through empirical causes, no, but because they were determined by grounds of reason" (A 550/ B 578).

difficult circumstances. But we still hold people accountable for their actions in such situations because we regard the *capacity* for reason and action according to moral maxims as paramount for freedom and not the actual manifestation of this capacity in particular circumstances. We may hold someone morally responsible in such cases because we recognize that she retains the capacity for reason regardless of whether the demands that reason makes accord with the possibilities of nature. And we hold someone morally responsible regardless of whether in fact she could have done otherwise given the state of the empirical world. What we regard as important is that we retain the capacity to act from freedom and moral imperatives not the particular ways in which such actions will be manifested in our empirical character and the empirical actions that flow from it. And Kant appears to think that we retain this capacity for reason so long as we retain our power of spontaneity.[215]

5.3 Natural Laws and The Capacity for Reason

It is often argued that Kant's account of freedom of the will requires the ability to change the laws of nature, which is a prospect that some commentators have found objectionable. Freedom for Kant requires the ability to have done otherwise. We can do otherwise in the sense that we could have manifested a different intelligible character and so also manifested different empirical actions. Philosophers have recognized however, that the "ability to have done otherwise" implies the ability to change the laws of nature.[216] Peter van Inwagen has formulated this objection in his so-called 'consequence argument'. According to the

---

[215] The fact that the capacity for reason or the capacity to break when thrown against a wall are timeless in this sense does not, however, mean that such a capacity exists eternally. A particular glass has the capacity to break when thrown against the wall if and only if the glass has certain causal powers. Similarly, a person retains the capacity to act according to reasons if and only if they retain the power of spontaneity that grounds our capacity for reason. There may, however, also appear to be a disanalogy between the capacity of the glass and the capacity for reason. In the case of reason, this capacity is not exhausted by its manifestation in a particular way in a given situation, whereas it may appear that the capacity for a glass to break is exhausted when the glass is broken. But just as a person can continue to act from reason so long as they possess the spontaneity that grounds this capacity, so too can a glass continue to have the capacity to break into ever smaller pieces.

[216] For an extensive discussion of this objection, see Tobias Rosefeldt, "Kants Kompatibilismus," in *Sind wir Bürger zweier Welten?: Freiheit und moralische Verantwortung im transzendentalen Idealismus*, ed. M. Brandhorst, A. Hahmann, B. Ludwig (Hamburg: Felix Meiner Verlag, 2012), pp. 91–96.

argument, if determinism is true, then all of our actions are the consequences of laws of nature and of previous events. Since it is not within our ability to change the laws of nature, the consequences of these laws, which include our present actions, are not up to us.[217] His argument is based on the intuitive idea that a person might be ordered to do something that is impossible given the laws of physics. If the laws of nature were up to us, then a person would be able to carry out such a task by simply wishing to do so, which appears absurd. We might, for example, demand that someone stop a bullet in mid air in order to save someone. And if the laws of nature were up to her, she could simply do so by wishing to, since she could make the laws such that they would allow her body to move fast enough to stop the bullet. Other arguments can also be made that show that the ability to change the natural laws is absurd since, for example, it would entail backward causation and the ability to alter the past. If Kant is committed to our ability to change the natural laws, and the consequence argument is correct, then it would appear that Kant has failed to demonstrate the compatibility of freedom and natural necessity.

Eric Watkins and others have attempted to answer this charge by arguing that Kant is able to show that the idea that we have the ability to change the laws of nature is coherent. On Watkins' account, Kant is committed not merely to the truth of the counterfactual, 'if her intelligible character had been different, the laws of nature would have been different', but to the idea that we can exercise certain causal powers that would make it the case that the laws of nature would be different. He defends this idea by arguing that for Kant empirical natural laws supervene on the empirical character, or nature, of substances in the empirical world, and the empirical character of substances supervenes on their intelligible character. So, the intelligible character of a substance determines its empirical character, which in turn determines what kinds of natural laws obtain in the empirical world.[218] Kant suggest such a view in the *Critique of Judgment*, for example, when he writes: "It is true that when we use the word *cause* with regard to the supersensible, we mean only the *basis* that determines natural things to exercise their causality to produce an effect in conformity with the natural laws proper to that causality" (AA 5:195–6), and in the *Critique of Pure Reason*, we have seen that Kant appears to suggest that reason must conform with natural necessity in the sense that it must determine it. From this account it becomes clear that if my manifestation of

---

[217] See Peter van Inwagen, *An Essay on Free Will* (New York: Oxford University Press, 1983), p. 56.
[218] See Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005), p.334.

some intelligible character had been different, then the natural laws that are instantiated would have been different. As Ben Vilhauer and others have argued, however, these natural laws must only be particular natural laws regarding individual substances. This is to say that the laws instantiated by an empirical character through the choice of an intelligible character must govern only the particular substance whose empirical character has been determined through the choice of intelligible character. This is the case because if the natural laws that are instantiated governed many substances, then the manifestation of one intelligible character might preclude the manifestation of another intelligible character. In the case of the free actions of persons, this would mean that my manifestation of my intelligible character might rob someone else of their freedom to manifest their intelligible character.[219]

Such an account is not without its difficulties. First, one might doubt that this account entails determinism about appearances. If at any moment the manifestation of some intelligible character rather than another can effect a change in the particular causal laws governing appearances and thus also the particular empirical actions that take place, it is unclear why appearances are thought to be deterministic. If empirical actions are grounded in the free activities of things in themselves, it appears to be an error to hold that determinism is true of appearances. Or at the very least, the appearance of determinism is the result of our inability to see that the interaction of empirical substances is grounded in freedom. Second, it is difficult to understand how the instantiation of particular natural laws through our manifestation of an intelligible character is compatible with Kant's explicit claim in the Third Analogy and elsewhere that causal determinism requires thoroughgoing causal interaction. He writes here that "all substances, insofar as they are simultaneous, stand in thoroughgoing community (i.e. interaction with one another)" (A 211).[220] This is to say that all empirical substances must be linked together in a causal network such that a change in one substance causes a change in all others however subtle. But if only particular laws are instantiated by our manifestation of some intelligible character, it is not clear how the various particular laws and the substances they govern can be linked together into an overall thoroughgoing causality.

One way to overcome the second problem may be to appeal to the idea that thoroughgoing causation is a regulative not a constitutive idea. It is a legitimate principle that

---

[219] See Ben Vilhauer, "Incompatibilism and Ontological Priority in Kant's Theory of Free Will," in *Rethinking Kant: Volume I*, ed. Pablo Muchnik (Newcastle upon Tyne: Cambridge Scholars Publishing, 2008), pp. 26–32.

[220] See also the alternate formulation of the principle in the B edition (B 256).

allows us to organize scientific knowledge, but it does not actually govern phenomena. But this is an unsatisfying response since Kant appears to be interested in providing a metaphysical account of freedom and causal interaction.[221] A second way one might overcome the objection regarding thoroughgoing causation is to argue that the manifestation of an intelligible character brings about the instantiation of empirical laws that are so particular that they include within them specifications about how all other empirical objects will be. For example, my manifestation of an intelligible character brings about an empirical character that is hedonistic such that I am causally determined to drink all the beer in my vicinity. And this causal law also necessitates that after I drink one beer I will get another one. And my acquiring another beer is causally responsible for someone else pouring it, which is causally responsible for someone else washing a new glass, which is responsible for someone else's order being delayed and so on. But if the particular laws are so specific, then it would seem that they will eventually necessitate actions by other persons. If they do so, then such persons are not free in their actions, which is contrary to the idea that persons act from freedom. Or if the person has the ability to intervene in the causal chain through their manifestation of some intelligible character, then they have the ability to change the causal laws that I have instantiated through my choice. But if this is true, then the causal laws I cause to be instantiated are actually limited in their scope and particularity by the laws instantiated through another person's manifestation of an intelligible character. And if they are so limited, it is unclear how anything like thoroughgoing causation among empirical appearances is possible. And if such thoroughgoing causation is not possible, then it is not true that empirical appearances are subject to causal determinism. So if Kant is committed to the idea that our intelligible character determines which particular natural laws are instantiated, then he has argued against causal determinism and compatibilism and for incompatibilist libertarianism.[222] It may turn out that these consequences are acceptable or

---

[221] Such a defense may ultimately lead to a fictionalist account of freedom of the will according to which we merely regard ourselves "as if" we are free. On the "as if" in Kant, see Hans Vaihinger, *The Philosophy of 'As If': A System of the Theoretical, Practical and Religious Fictions of Mankind*, trans. C. K. Ogden (London: Kegan Paul & Company, 1924).

[222] On the debate about whether Kant is a compatibilist or incompatibilist, see: Simon Shengjian Xie, "What Is Kant: A Compatibilist or an Incompatibilist? A New Interpretation of Kant's Solution to the Free Will Problem," *Kant-Studien* 100 (2009), pp. 53–76; Ben Vilhauer, "Incompatibilism and Ontological Priority in Kant's Theory of Free Will," in *Rethinking Kant: Volume I*, ed. Pablo Muchnik (Newcastle upon Tyne: Cambridge Scholars Publishing, 2008), pp. 22–47; Ben Vilhauer, "Can We Interpret Kant as a Compatibilist

that the objections can be answered, but such problems do suggest that there are reasons for seeking an alternative interpretation of Kant's compatibilism that does not require that we have the ability to change the natural laws.

An alternate way of providing an account of compatibilism, which appears much closer to Kant's view and fits with his idea that reason is an intelligible capacity to act from maxims, may be found in a dispositionalist view of free will, according to which we have free will when we have certain abilities for acting or not acting. One finds an early version of this view in Hume's *Enquiry Concerning Human Understanding* (1748), where he writes: "By liberty, then, we can only mean a power of acting or not acting, according to the determinations of the will; this is, if we choose to remain at rest, we may; if we choose to move, we also may."[223] Similarly, in the German tradition with which Kant was familiar, freedom of the will is often spoken of as a power, ability, or faculty. Tetens, for example, argues that freedom requires a certain capacity to do otherwise, which is grounded in a spontaneous *Selbstmacht* that allows the soul to determine its actions.[224] And Wolff calls freedom a "faculty of the soul to decide."[225] We have also seen that freedom for Kant is characterized as a capacity or ability to do otherwise and that this capacity to do otherwise rests in the capacity of reason to give itself a moral maxim independent of antecedently determining conditions. The unrestricted manifestation of such an ability suggests a

---

About Determinism and Moral Responsibility?" *British Journal for the History of Philosophy* 12(4) (2004), pp. 719–730.

[223] See David Hume, *An Enquiry Concerning Human Understanding*, ed. Tom L. Beauchamp (Oxford: Oxford University Press, 2000), 8.1, p. 72, SBN 95.

[224] See Johann Nicolas Tetens, *Philosophische Versuche über die mensliche Natur und ihre Entwicklung*, vol. 2 (Leipzig: 1777), pp. 18, 20–21, 125. For a detailed discussion of Tetens' view on freedom of the will, see Andree Hahmann, "Tetens über die Freiheit als Vermögen der Seele" (forthcoming).

[225] See Christian Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (1720) (Halle: 1751), §519, where Wolff discusses freedom as a "faculty of the soul to decide," and Baumgarten, *Metaphysica* (Frankfurt: 1757), §719. One central point of disagreement in the period among those who held that freedom is a capacity is about whether the capacity for freedom is identical with or somehow grounded the capacity for reason. Both Crusius and Tetens argue against equating freedom with reason, whereas Wolff argues that freedom, as with any capacity, is reducible to a single fundamental power of the soul, which he equates with the soul's ability to represent. See Max Wundt, *Kant als Metaphysiker–Ein Beitrag zur Geschichte der deutschen Philosophie im 18. Jahrhundert* (Stuttgart: Ferdinand Enke, 1924) (reprinted Hildesheim: Olms 1984), pp. 62–64; Christian Wolff, *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* (1720) (Halle: 1751), §520). For Kant, reason exhibits freedom and spontaneity.

libertarian view of freedom according to which we simply have the capacity to begin an action without being causally determined by previous actions or events. However, the idea that freedom of the will is grounded in certain capacities that are grounded in our powers can also be given a compatibilist interpretation.

One way to analyze a capacity or ability is in terms of a conditional statement. For example, Kant has the ability to take a different path to the university if and only if in some circumstance he would take a different path if he were to try. But as Michael Fara has argued, this characterization of an ability runs into problems when one considers certain counter-cases in which an ability is masked. Kant's ability to take a different path in some circumstance is masked if the circumstance obtains but Kant fails to take a different path.[226] Thus Kant might try to deviate from this path, but he has forgotten how to get to the university by any other path. So when he strikes out on the novel path, he inadvertently follows the same path to the university. It turns out that the conditional that Kant has the ability to take a different path to work if he were to try is false. Nevertheless it might be argued that Kant retains the ability to take a different path to the university, although in some particular circumstance this ability is masked by various factors. Such scenarios have led Fara and others to argue that a better account of our abilities is one that holds that we retain our abilities so long as we retain the causal powers that ground these abilities.[227] In order to assess whether someone in some circumstance had the ability to do otherwise, and thus whether they are free, we hold these causal powers or intrinsic properties constant and then consider the range of possible worlds in which they retain these powers and these powers act in a way that is not hindered by other factors. For example, in a wide range of possible worlds, Kant retains his causal powers of motion and navigation and takes a different path to the university. This is so even if in a lot of other possible worlds, or counterfactual scenarios, Kant's ability to do this was masked or hindered. In some possible world, for example, Kant was compelled by some extrinsic factors to take the same path to the university each day.

---

[226] See Michael Fara, "Masked Abilities and Compatibilism," *Mind* 117(468) (2008), p. 850. For other dispositionalist accounts of freedom of the will, see Kadri Vihvelin, "Free Will Demystified: A Dispositional Account" *Philosophical Topics* 32(1/2) (2004), pp. 427–450, who argues that "to have free will is to have the ability to make choices on the basis of reasons and to have this ability is to have a bundle of dispositions" (429). For arguments against "new dispositionalism," see Randolph Clarke, "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind* 118 (470) (2009), pp. 323–351, and Ann Whittle, "Dispositional Abilities," *Philosophers' Imprint* 10(12) (2010), pp. 1–22

Nevertheless, across all of these possible worlds, Kant retains the causal powers that would have enabled him to do otherwise had he so chosen. So, Kant retains the ability to do otherwise regardless of the outcome of these various scenarios. And if freedom of the will is understood as the ability to do otherwise, then Kant has freedom of the will in each of these scenarios.

It may help to illustrate the relevance of this view for Kant's theory of freedom by returning to his example of the malicious lie. Kant argues that we have the power of spontaneity required to ground our capacity for reason, which is a capacity to act according to grounds that are not sensible and therefore not determined by antecedent events and actions. Now consider the scenario in which a person retains the capacity to act according to the moral maxims prescribed by reason, but they fail to tell the truth because unbeknownst to them their brain has been manipulated in such a way that they will always say the opposite of what they intend to say. Despite the fact that their capacity for reason manifests itself in such a way that leads to a lie in the empirical world, they nevertheless retain the ability to do otherwise. They retain the ability to do otherwise because they retain the capacity for reason. And they meet the necessary conditions for moral responsibility for their actions because they retain this capacity for reason. Now, it may also be the case that the liar tells the truth only in those possible worlds in which the natural laws are different or the series of events leading to the lie are different. But the alteration of these natural laws is not important for assessing whether the liar has the ability to do otherwise. The liar may not have the ability to alter the natural laws because she lacks the powers necessary for such abilities, but she nevertheless retains the ability to act according to the capacity for reason and so also the ability to do otherwise.[228] This ability just fails to be manifested in certain circumstances.

If we interpret Kant's account of freedom and our ability to do otherwise in this way, then it is unnecessary to show that we have the capacity to change the natural laws, as Watkins and others attempt to do, because the ability to change the natural laws is not necessary for our ability to do otherwise. We are able to do otherwise and are free in this sense when we retain certain causal powers, particularly the power of spontaneity, and the capacity for reason that could have been manifested had we chosen to manifest it. Although this capacity was not actually manifested in some circumstance, and even if something would have intervened to prevent the manifestation of this capacity, one nevertheless acted freely

---

[228] My reconstruction here is indebted to Ann Whittle, "Dispositional Abilities," *Philosophers' Imprint* 10(12) (2010), p. 16.

because one retained the power that would have enabled one to do otherwise. Such an account is also compatible with the natural-laws account provided by Watkins and others because it may be that in some cases we do have the powers and abilities necessary to change the natural laws. But this ability is not necessary for the ability to have done otherwise. So, many of the worries raised for the natural-laws account may be resolved if we dispense with the idea that Kant held that our freedom in the sense of our ability to do otherwise entails an ability to change the natural laws that obtain. Rather, our ability to do otherwise simply consists in our continuing possession of the power of spontaneity that grounds our capacity for reason.

Although this account may provide some support to indicate that Kant need not be committed to our ability to alter the natural laws when he suggests that freedom requires the ability to do otherwise, one might nevertheless wonder how Kant's view differs from contemporary dispositionalist accounts. Kant's view clearly differs in the details, particularly with respect to the analysis of our abilities in terms of possible worlds. Nevertheless, Kant has the important insight that our freedom depends on our possession of certain intelligible capacities and powers rather than our manifestation of these capacities in the empirical world. As I pointed out previously, this is one reason Kant considers why we tend to hold that someone had the ability to do otherwise even in cases in which the demands made by reason do not accord with what is possible according to natural laws. Importantly, Kant also recognizes that freedom cannot consist in the manifestation of these capacities as actions in the empirical world, since these actions would be subject to causal determinism. Rather freedom must consist in the capacity or ability to act according to reason and the moral law whether one actually does so or not.

5.4 Moral Responsibility and The Capacity for Reason

One problem that may be raised for the preceding interpretation of Kant's views on freedom and our ability to do otherwise is that it may lead to counterintuitive results regarding moral responsibility. One might consider the kind of case proposed by Harry Frankfurt in which one does not have the local ability to act otherwise, but one nevertheless retains a global ability to act otherwise.[229] For example, a person is faced with the choice of lying or telling the truth in

[229] Frankfurt-style cases are intended as counterexamples to the principle of alternative possibilities, which holds that someone is morally responsible for some action only if they

some situation. Unbeknownst to this person, however, an evil scientist has gained control over the person's mind and motor functions. And if the person should decide to tell the truth, the evil scientist will intervene to force the person to lie. In this case, the person does not have the local ability to do otherwise than to lie. They nevertheless retain the causal powers required for telling the truth and therefore also the global ability to tell the truth if they wished to, although this ability is masked in this particular situation. One might think that on the interpretation proposed above the person may be counted as morally responsible for their malicious lie because they retain such a capacity. But such cases are not worrisome for Kant's account as I have interpreted it. First, it is not certain whether one should not argue in such situations that the global ability to do otherwise is sufficient for freedom of the will and also sufficient for moral responsibility. As Kant points out, we sometimes do hold persons morally responsible even when the imperatives given by reason do not accord with what is possible in nature. More importantly, however, Kant's primary aim in his discussion of freedom of the will surrounding the Third Antinomy is merely to argue that freedom of the will is conceivable despite determinism and that such freedom is a necessary condition for moral responsibility. He does not aim to argue that the kind of freedom he describes is sufficient for moral responsibility. It remains possible that there are other factors that need to be considered when assessing moral responsibility, such as whether one's ability to do otherwise was in fact in accord with what is possible in nature.

Some interpreters have also worried that Kant's account of freedom of the will seems to be committed to the view that all noumenal beings, not only noumenal persons, possess freedom of the will.[230] And if it is true that all noumenal beings possess freedom, then we would have to conclude that we should hold stones, for example, as morally responsible for their actions as persons. But there is little evidence that Kant commits himself to this view. Kant considers this question in the *Prolegomena*, where he argues that "we cannot bestow freedom upon matter in consideration of the unceasing activity by which it fills its space, even though this activity occurs through an inner principle" (AA 4:344). It is clear in the *Critique of Pure Reason* that matter lacks freedom because it is not within the causal powers

---

could have done otherwise. See Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66(23) (1969), pp. 829–839.

[230] For a variation on this line of criticism, see Jonathan Bennett, "Kant's Theory of Freedom," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), p. 105. Bennett is wrong to deny that Kant maintains that only entities with the capacity for reason have noumenal freedom.

or intrinsic properties of matter to have an intellectual capacity, a capacity for reason. Kant is quite explicit that freedom of the will requires the capacity for reason. And there is no reason to think that stones posses this capacity. Nor does Kant seem to think that animals and children possess such a capacity.[231] Moreover, it does not appear that Kant would think that such entities should be held morally responsible for their lack of this capacity. These capacities are grounded in the causal powers and intrinsic properties of things as they are in themselves. And we do not have the freedom to choose the causal powers that ground our capacities. Because we have no freedom to choose which causal powers we have, we are not morally responsible for having or not having these causal powers and therefore not responsible for the capacities we have or do not have. Someone who cannot walk is not morally responsible for failing to choose to have the power to walk since they have no such choice. Likewise someone who is incapable of acting according to the moral law because of cognitive deficiencies is not morally responsible for failing to choose to have the powers that ground the capacity for reason. And the same is true of animals and objects. However, an entity that does possess the causal powers that ground the capacity for reason meets the necessary conditions for moral responsibility.

But even if the kinds of entities that possess freedom of the will can be restricted to persons, one might still worry about the implications of Kant's account as I have interpreted it for determining the extent of our moral responsibility for actions. Ralph Walker has argued that if a person's intelligible character determines their empirical character and thereby also the empirical actions that flow from this character, then the effects of their manifestation of this character will extend indefinitely because all empirical actions and events are in causal interaction. The consequence is that I may be held morally responsible for all actions in the world, including those of other agents: "I can be blamed for the First World War, and for the Lisbon earthquake that so appalled Voltaire. Gandhi is not less guilty than Amin of the atrocities of the Ugandan dictator."[232] Allen Wood has attempted to answer this worry by

---

[231] Kant writes: "In the case of the lifeless nature and nature having merely animal life, we find no ground for thinking of any faculty which is other than sensibly conditioned. Yet the human being, who is otherwise acquainted with the whole of nature solely through sense, knows himself also through pure apperception, and indeed in actions and inner determinations which cannot be accounted at all among impressions of sense" (A 546f./ B 574f.).

[232] See Ralph C.S. Walker, *Kant* (London: Routledge & Kegan Paul, 1978), p. 149. For discussions of the extent of our moral responsibility in Kant's theory of freedom, see also: Wolfgang Ertl, "Schöpfung und Freiheit: Ein kosmologischer Schlüssel zu Kants

arguing that since we cannot know with certainty for which actions we are morally responsible it is open to Kant to suppose that "they correspond to those events for which we normally regard ourselves as morally responsible."[233] But this response is deeply unsatisfying because it simply claims ignorance of the extent of our moral responsibility and proposes that our attributions of praise and blame should remain just as they are in the face of this ignorance. Interpretations that attempt to extricate Kant from this problem by arguing that one is morally responsible only for those actions that follow from the laws that one instantiates through one's manifestation of an intelligible character are also unsatisfying for the reasons mentioned previously.[234]

However, one might instead simply argue against Walker that the problem he raises is interesting but is not a genuine problem for Kant. It is clear that on Kant's account we are responsible for our manifestation of an intelligible character. But whether this manifestation and the accompanying empirical acts issue in unforeseen consequences or consequences that are beyond a person's power to change because of the deterministic nature of appearances is not something for which the person may be held morally responsible. Kant believes that when judging the moral value of an action we should focus on the intentions rather than the consequences of an action. Thus whether my manifestation of an intelligible character that obeys the dictates of reason and the moral law issues in some undesirable action because of how this action interacts with others within the empirical world of appearances does not matter for the moral value of my intellectual character. A person is morally praiseworthy so long as they act in accord with the moral maxims given by reason. And they are morally neutral with respect to the empirical consequences of the manifestation of their capacity for

---

Kompatibilismus," in *Kants Metaphysik und Religionsphilosophie* (Hamburg: Felix Meiner, 2004), p. 65; Jonathan Bennett, "Kant's Theory of Freedom," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), p. 105; Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984).

[233] Allen Wood, "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), p. 92.

[234] See Ben Vilhauer, "Incompatibilism and Ontological Priority in Kant's Theory of Free Will," in *Rethinking Kant: Volume I*, ed. Pablo Muchnik (Newcastle upon Tyne: Cambridge Scholars Publishing, 2008). For a similar view, see Robert Hanna and A. Moore, "Reason, Freedom and Kant: An Exchange" *Kantian Review* 12(1) (2007), p. 121. See also Vilhauer, "Can We Interpret Kant as a Compatibilist About Determinism and Moral Responsibility?" *British Journal for the History of Philosophy* 12(4) (2004), pp. 719–730; "The Scope of Responsibility in Kant's Theory of Free Will," *British Journal for the History of Philosophy* 18(1) (2010), pp. 45–71.

reason. Although a person may be the ultimate source of some action when we trace this action through the causal chains, they may not be morally praiseworthy or blameworthy for the action. Moral praise and blame have much more to do with one's intelligible moral character and its associated maxims than with the actions that are manifested through this character. However, this kind of response also raises a number of additional questions.

One objection that may be raised is this. How, on the interpretation provide here, can one distinguish between a good and a bad intelligible character and thus assign moral praise or blame to this intelligible character? On the interpretation provided here, freedom as the ability to do otherwise consists merely in possessing the capacity for reason grounded in our spontaneity that allows one to formulate and act according to the moral law. But it seems that all persons according to Kant possess this capacity. A liar and an honest person possess this capacity for reason in equal degree. The difference between the liar and the honest person, however, is that the honest person acts according to the moral law given by reason whereas the liar does not.[235]

One response here is just to argue that the interpretation thus far has focused primarily on understanding what freedom as the ability to do otherwise might mean for Kant. On this account, the liar and the honest person are both free insofar as they possess the capacity for reason and the power of spontaneity. And this kind of freedom is a necessary condition for moral responsibility, so the liar and the honest person both meet the necessary condition for moral responsibility. In asking about the good or bad moral character to which moral praise and blame are supposed to attach, however, we are wondering about a sufficient condition for moral responsibility. This is to say that if we can determine whether some intelligible character is good or bad or followed the right moral maxims or not, then this would be sufficient for holding them morally responsible regardless of the ultimate outcomes of their actions in the empirical world. If someone has a good character and acted from good intentions, then this is sufficient for thinking they are morally laudable. However, Kant suggests that we can never know what the intelligible character of a person is but only how this character is manifested in an empirical character: "The empirical character is once again determined in the intelligible character […]. We are not acquainted with the latter, but it is indicated through appearances, which really give only the mode of sense (the empirical character) for immediate cognition" (A 551/ B 579). And so it seems that we can never genuinely know whether someone is worthy of praise or blame. As Kant writes in a note

---

[235] Thanks to Rachel Zuckert for raising this and the subsequent objection.

appended to the previous passage: "The real morality of actions (their merit and guilt), even that of our own conduct, therefore remains entirely hidden from us. Our imputations can be referred only to the empirical character. How much of it is to be ascribed to mere nature and innocent defects of temperament or to its happy constitution (*merito fortunae*) this no one can discover, and hence no once can judge it with complete justice" (A 551/ B 579).

Problematically, however, even if this response is granted, the entire discussion of moral accountability seems to suggest that one has some choice of one's intelligible character, and this choice of intelligible character requires timeless agency. If someone is to be praised or blamed for their intelligible character, then it seems necessary that they have some choice in their intelligible character. And this choice must be timeless for the reasons Kant suggests. So it seems that the timeless choice of an intelligible character is necessary for moral accountability. In response, the following might be argued. One's intelligible character possesses a disposition to act in a certain way. This disposition is timeless in the way that any other disposition is timeless as was shown above. An intelligible self cannot, however, choose between two dispositions, i.e. intelligible characters. In the case of the liar, the intelligible self cannot choose to have the disposition to lie or to tell the truth. However, the intelligible self has the power to manifest this disposition or character or not. This is to say that if the intelligible self is disposed to lie, it may not choose to tell the truth, but it may refrain from lying. This is to say that the power in one's intelligible self is only a power to refrain or block the manifestation of some disposition. We can put the brakes on the disposition we are given, but we are not free to choose this disposition. Even in our intelligible characters, we are not morally blank slates. And accordingly praise and blame do not attach to the dispositions of our intelligible self but only to the intelligible manifestation of some disposition. But this kind of response is unlikely to be satisfying since there still remains the sneaking suspicion that the power to manifest or refrain from manifesting some disposition is nevertheless a choice. And if no sense can be made of the idea of timeless agency, then it is a choice in time. And if it is a choice in time, then it appears that it is subject to causal determinism and we are not, after all, free in our intelligible character. One might only speculate that such problems partially contribute to Kant's abandonment of the account of freedom of the will in the *Critique of Pure Reason* in favor of another line of argument in the *Groundwork* and the *Critique of Practical Reason*.

Before concluding, it may also be worthwhile to consider another worry that may be raised for this interpretation of Kant's account of the freedom of the will. On Kant's account, freedom is supposed to be a kind of causality that differs from the causality involved in

empirical situations. Whereas empirical causality presumes thoroughgoing causal determination, intelligible causality is the freedom to begin a series on its own.[236] And indeed the core of the Third Antinomy is dedicated to understanding the sense in which these two kinds of causality might be compatible. So it might be wondered how intelligible causality fits in to the preceding interpretation of Kant's account of freedom of the will. There are a few responses here.

One response is simply to argue that Kant's account of intelligible causality as the ability to begin a series on one's own without any reference to other conditions is hopeless. And if freedom requires this kind of causality, then his account of freedom of the will is hopeless. In one regard, the interpretation I have provided can, however, provide a weaker conception of intelligible causality. Insofar as the interpretation has focused on Kant's characterization of freedom as the ability to have done otherwise, it does not require an account of intelligible causality as the ability to begin a causal series without reference to any other conditions. We have the ability to do otherwise and thus have freedom of the will when we possess the capacity for reason and the powers that make such reason possible. This does not, however, require that we have any absolute ability to begin a causal series on our own without respect to other conditions. Indeed, on the account I have provided, even our intelligible causality is conditioned in some regard. In order for it to be effective, it must conform to the world of appearances and the given laws of nature. This means that we can initiate actions through our intelligible character and intelligible causality but these actions come to fruition only when they conform to the world of causally determined empirical appearances. One might think of this view using a picture. Intelligible causality is sort of like a car that wishes to merge onto an interstate with the other traffic. The other traffic is the empirical realm. Intelligible causality is effective in the empirical realm when it somehow fits with laws of the empirical realm. In the analogy, the car is able to merge with the traffic because there are gaps that allow it to do so. In other cases, the natural laws block the manifestation of the intelligible cause. In the analogy, the merging car cannot find any way into the traffic. What is important for Kant's account of freedom of the will, however, is not whether our intelligible causality is effective in an unconditioned way but whether we continue to possess the capacity of reason and the powers that ground it that allow that we could have done otherwise.

One might wonder, however, about the degree to which this reconstructive interpretation respects Kant's view. It clearly does not accord with moments in which Kant is talking about

---

[236] See A 538/ B 566 – A 539/ B 567.

an intelligible cause as an ability to begin a series on its own regardless of any conditions. As when he says: "By freedom […] I understand the faculty of beginning a state from itself, the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature" (A 533/ B 561). It does, however, fit well with passages pointed to earlier in which Kant suggests that our intelligible causality must fit somehow with the natural laws and suggests that this causality is sometimes blocked. As when he says "the action [demanded by reason] must be possible under natural conditions if the ought is directed to it; but these natural conditions do not concern the determination of the power of choice itself, but only its effect and result in appearance" (A 547f./ B 575f.). This is to say that the fact that one possesses intelligible causality as the ability to begin a series on one's own does not mean that this intelligible causality can be exercised without restriction and without conformity with natural necessity. The virtue of the view, however, is to have overcome to some degree Kant's reliance on the seemingly incoherent idea of timeless agency and the ability to change the laws of nature while nevertheless providing a metaphysical interpretation of Kant's account of freedom of the will and its compatibility with causal determinism.[237]

5.5 Conclusion

In the preceding discussion, I have argued that Kant provides a solution to the problem of freedom of the will that shows that freedom of the will is possible and that it is compatible with causal determinism. On Kant's account, we have a capacity for reason, which is the capacity to act in accordance with moral maxims. This capacity is grounded in our power of spontaneity. Although reason demands that the moral maxims and actions that flow from these maxims should accord with the possibilities of nature and the natural laws, it is often the case that reason makes demands that cannot be met given empirical nature. In such cases, reason demands an action, but the exercise of this action is somehow masked by deterministic causal events such that a person fails to carry out an action that was based on a particular

[237] This interpretation also resolves the problem of causal overdetermination. Kant suggests at times that empirical causality is not sufficient but empirical actions also require intelligible causality. See, for example, A 446/ B 474. But this seems to suggest that empirical actions are causally overdetermined since they have both an empirical and an intelligible cause. On the interpretation I have argued for intelligible causality is, however, only effective when it conforms with the natural necessity of empirical appearances, so there appears to be no causal overdetermination of empirical actions.

moral maxim. In such cases, however, one nevertheless retains the capacity for reason and therefore also the ability to do otherwise than one has done regardless of what happens in the empirical world. For Kant, the retention of this capacity is sufficient for freedom regardless of how such freedom is manifested in the empirical world. On the basis of this conception of freedom, I have also provided answers to some problems regarding timeless agency, natural laws, and the extent of moral responsibility that plague metaphysical interpretations of Kant's account of freedom. The notion of timeless agency can be given a coherent interpretation if one recognizes that the capacity for reason is timeless in the sense that one retains this capacity even when it is not manifested in actual situations. Kant's account need not be committed to the idea that the ability to do otherwise entails the ability to change the laws of nature. We retain the ability to do otherwise when we retain the capacity for reason, which is grounded in our power of spontaneity, regardless of whether we have the kinds of causal powers that would allow us to change the laws of nature. Kant also restricts moral responsibility to only those things that possess the capacity for reason and the power of spontaneity that ground this capacity. And he argues that a person is morally responsible only for their exercise of the capacity for reason and not for the outcomes of the exercise of this capacity in empirical actions.

Conclusion

We began by considering the legacy of German rationalist metaphysics in Kant's discussions of the metaphysics of mind. We saw that although Kant criticizes the epistemic warrant for rationalist claims regarding the substantiality, simplicity, personality, immortality, and freedom of the soul, he nevertheless retains some of their metaphysical insights and seeks to accommodate them or transform them within his transcendental idealism. In his confrontation with the basic tenets of rational psychology regarding the substantiality and simplicity of the soul, Kant argues that if there is an ultimate ground of thought, then it must be a thing it in itself, and that if it is a substance it cannot be a persisting substance but must only be thought of as something that grounds the attributes of thought through its intrinsic powers. Kant also shows that the possession of multiple mental powers that work together jointly is necessary for cognition and the unity of thought and that the existence of a multiplicity of powers does not entail that the ground of thought is a composite substance. Kant's account of the ground of thought as a substance endowed with powers also supports his account of personhood, the possibility of mind-body interaction, and the possibility of freedom of the will. Personhood for Kant is retained so long as the fundamental powers that ground our capacity for unified thought are retained. Mind-body interaction can occur through physical influx because it is possible that the substances that ground mental and physical appearances are intrinsically neither mental nor physical and may causally affect one another through their powers. And freedom of the will is possible and reconcilable with causal determinism because we have a capacity for reason that allows us to provide ourselves a moral maxim independent of antecedent causes, and we retain this capacity regardless of how it may be manifested in the empirical world. On the whole, we have seen that Kant does not wholly reject rationalist metaphysics on the basis of his critical epistemology and that many of his positions on the substance that grounds thought share important similarities with the views of his predecessors and some of his positions in his pre-critical writings.

We have also seen that one of the motivations for Kant retaining a number of positions from his rationalist predecessors and altering them to fit within the doctrine of transcendental idealism is in part to provide an account of the necessary conditions for moral responsibility. Such an account is important not only for the theoretical concerns of the *Critique of Pure Reason* (1781/1787) but also provides a foundation for Kant's later accounts

of moral responsibility and practical philosophy in the *Groundwork of the Metaphysics of Morals* (1785) and the *Critique of Practical Reason* (1788). Thus in the *Critique of Pure Reason*, Kant provides an account of the necessary conditions for personhood because the retention of personhood is a necessary condition for moral responsibility. He also argues that freedom of the will is necessary for moral responsibility, and he shows how such freedom is possible in light of causal determinism. Importantly, Kant is also able to establish the possibility of these features required for moral responsibility without being guilty of the metaphysical hubris he accuses the rationalist metaphysicians of in their accounts of the nature of the soul and the features of the soul necessary for moral responsibility.

This account of Kant's metaphysics of mind and its relationship to rational psychology also suggests that a dominant interpretative position, which maintains Kant's aim in the Paralogisms and elsewhere is primarily to criticize the epistemic warrant for rationalist claims regarding the nature of the soul, only tells part of the story, since there is ample evidence in the *Critique of Pure Reason*, Kant's lectures on metaphysics, and his *Reflexionen* that shows Kant not only struggled with his own indebtedness to his rationalist predecessors and his pre-critical rationalist positions but that many of his commitments to rationalist positions are central to his views on the metaphysics of mind and his practical philosophy. Understanding Kant along these lines also reveals that there is a much tighter relationship between his earlier metaphysical interests and his critical philosophy than previously thought.

Having shown that Kant provides a conception of the metaphysics of mind that establishes certain features of the mind that are necessary for practical philosophy and that he secures this conception of the mind through his confrontation with rationalist metaphysics, there are nevertheless a number of questions that remain open. In the second edition of the *Critique of Pure Reason*, Kant appears to provide a fundamental revision of his account of thought and its ontological ground. And it has often been argued that Kant undertook these revisions because he recognized that many of the claims regarding the nature of things in themselves in the first edition could not be supported given the epistemic restrictions provided by the doctrine of transcendental idealism. Thus it is thought that in the second edition of the Paralogisms Kant rejects a substantial account of the ground of thought in favor of a view that maintains that the determining self is systematically elusive and not a possible object of knowledge. But it is unclear whether Kant's revisions in the second edition are actually so drastic and whether they might not be compatible with the more metaphysical and rationalist formulations in the first edition. And even if Kant's revisions are as drastic as sometimes thought, it also remains to be seen whether the revisions in the second edition are

able to provide an account of persons, mind-body interaction, and freedom that appear to be required for moral responsibility and thus whether Kant was right to have revised his views.

One might also wonder why Kant abandons the argument for freedom of the will in the *Critique of Pure Reason* in favor of a reformulation of the argument in the *Groundwork of the Metaphysics of Morals* and the *Critique of Pure Practical Reason*. Neither argument appeals to the notion of intelligible causality or timeless agency found in the first *Critique*, and so one might think that Kant abandoned the picture of substance causality at work in his conception of metaphysics of mind in the first *Critique* in favor of some other view. In these latter works on practical philosophy, Kant also appears to attempt to recuperate the notion of an immortal soul in order to explain the conditions under which we may strive for moral perfection in our ethical lives without adopting the rationalist view that the existence of an immaterial substantial ground of thought entails that it is immortal.

The discussion of personhood and its relationship to moral responsibility in Kant also raises questions about the role of Kant's metaphysics of mind in his practical philosophy as a whole. It was argued that certain mental capacities are required for personhood and that these are the mental capacities that are also required for rational cognition and for deliberation about moral action. But there is still much more to be said about how these powers contribute to rational cognition on Kant's account of the mind and cognition and how they also support the cognition required for rational deliberation. However, it is perhaps this conception of personhood filled out by the account of mental causation and freedom of the will that provides a clue to understanding how Kant transformed the rationalist doctrine of the soul and mental powers into a conception of moral agents in his practical philosophy.

Bibliography

Adickes, Erich. *Kant und das Ding an Sich*. Berlin: Pan Verlag Rolf Heisse, 1924.

Allais, Lucy. "Kant's One World." *British Journal for the History of Philosophy* 12(4) (2004): 655–684.

———. "Kant's Transcendental Idealism and Contemporary Anti-realism." *International Journal of Philosophical Studies* 11(4) (2003): 369–392.

———. "Intrinsic Natures: A Critique of Langton on Kant." *Philosophy and Phenomenological Research* 73(1) (2006): 143–169.

———. "Kant's Idealism and the Secondary Quality Analogy." *Journal of the History of Philosophy* 45(3) (2007): 459–484.

Allison, Henry. *The Kant–Eberhard Controversy*. Baltimore: The Johns Hopkins University Press, 1973.

———. *Kant's Transcendental Idealism. An Interpretation and Defense*. New Haven: Yale University Press, 1983.

———. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.

———. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. Cambridge: Cambridge University Press, 1996.

———. *Kant's Transcendental Idealism. An Interpretation and Defense*. Revised and enlarged edition. New Haven: Yale University Press, 2004.

———. "Transcendental Realism, Empirical Realism and Transcendental Idealism." *Kantian Review* 11 (2006): 1–27.

Ameriks, Karl. "Recent Work on Kant's Theoretical Philosophy." *American Philosophical Quarterly* 19(1) (1982): 1-24.

———. "The Critique of Metaphysics: Kant and Traditional Ontology." In *The Cambridge Companion to Kant*, edited by Paul Guyer, 249–279. Cambridge: Cambridge University Press, 1992.

———. *Kant's Theory of Mind. An Analysis of the Paralogisms of Pure Reason*. 2nd Edition. New York: Oxford University Press, 2000.

———. *Interpreting Kant's Critiques*. Oxford: Oxford University Press, 2003.

———. "Kant's Deduction of Freedom and Morality. In *Interpreting Kant's Critiques*, 161–192. Oxford: Oxford University Press, 2003.

———. “Kant’s *Groundwork* III Argument Reconsidered. In *Interpreting Kant’s Critiques*, 226–248. Oxford: Oxford University Press, 2003.

———. *Kant and the Historical Turn: Philosophy as Critical Interpretation*. Oxford: Oxford University Press, 2006.

———. “Kantian Apperception and the Non-Cartesian Subject.” In *Kant and the Historical Turn: Philosophy as Critical Interpretation*, 51–66. Oxford: Oxford University Press, 2006.

Atherton, Margaret. “Locke’s Theory of Personal Identity.” *Midwest Studies in Philosophy* 8 (1) (1983): 273–293.

Ayers, Michael. “The Ideas of Power and Substance in Locke’s Philosophy.” *Philosophical Quarterly* 25 (1975): 1–27.

———. *Locke*. Vol.2, Part I. London: Routledge, 1991.

Baumgarten, Alexander Gottlieb. *Metaphysica*. 1739. Frankfurt: 1757.

———. *Metaphysik*. Translated by Georg Friedrich Meier. Halle: 1766.

———. *Metaphysica*. 1739. Edited by G. Gawlick and L. Kreimendahl. Stuttgart-Bad: Fromann-Holzboog Verlag, 2011.

Beck, Lewis White. *Early German Philosophy*. Cambridge: Harvard University Press, 1969.

Bennett, Jonathan. “Kant’s Theory of Freedom.” In *Self and Nature in Kant’s Philosophy*, edited by Allen Wood. Ithaca: Cornell University Press, 1984.

———. *Kant’s Analytic*. Cambridge: Cambridge University Press, 1966.

———. *Locke, Berkeley, Hume*. Oxford University Press, Oxford, 1971.

———. *Kant’s Dialectic*. Cambridge: Cambridge University Press, 1974.

———. “Substratum.” *History of Philosophy Quarterly* 4 (1987): 197–215.

Bennett, Karen. “Mental Causation.” *Philosophy Compass* 2/2 (2007): 316–337.

Bermudez, Jose Luis. “The Unity of Apperception in the *Critique of Pure Reason*.” *European Journal of Philosophy* 2 (1994): 213–240.

Bilfinger, Georg Bernhard. *Dilucidationes philosophicae*. Tübingen: 1746.

Bird, Graham. *Kant’s Theory of Knowledge. An Outline of One Central Argument in the ‘Critique of Pure Reason’*. London: Routledge & Kegan Paul, 1962.

———. “The Paralogisms and Kant’s Account of Psychology.” *Kant-Studien* 91(2) (2000): 129-145.

———. editor. *A Companion to Kant.* Malden: Blackwell Publishing, 2006.

Blackwell, Richard J. "Christian Wolff's Doctrine of the Soul." *Journal of the History of Ideas* 22(3) (1961): 339–354.

Boehm, Omri. "The Principle of Sufficient Reason, the Ontological Argument and the Is/Ought Distinction." *The European Journal of Philosophy*. Forthcoming.

Broad, C.D. *Kant: An Introduction*. Cambridge: Cambridge University Press, 1978.

———. *The Mind and Its Place in Nature*. 1925. Oxon: Routledge, 2001.

Brook, Andrew. *Kant and the Mind*. Cambridge: Cambridge University Press, 1994.

Buroker, Jill Vance. *Kant's Critique of Pure Reason: An Introduction*, Cambridge: Cambridge University Press, 2006.

Caranti, Luigi. "Kant's Criticism of Descartes in the "Reflexionen zum Idealismus" (1788–1793)." *Kant-Studien* 97 (2006): 318–342.

———. *Kant and the Scandal of Philosophy*. Toronto: Toronto University Press, 2007.

Carl, Wolfgang. *Der schweigende Kant*. Göttingen: Vandenhoeck & Ruprecht, 1989.

———. "Kant's Refutation of Problematic Idealism: Kantian Arguments and Kant's Argument against Skepticism." In *A Companion to Kant*, edited by Graham Bird, 182–191. Malden: Blackwell, 2010.

Carpenter, Andrew N. *Kant's Earliest Solution to the Mind/Body Problem*. PhD dissertation, University of California, Berkeley, 1998.

———. "Kant's First Solution to the Mind/Body Problem." In *Kant und die Berliner Aufklärung*, edited by V. Gerhardt, R. Horstmann, & R. Schumacher, volume 2, 3–12. Berlin: De Gruyter, 2001.

Cassam, Quassim. "Kant and Reductionism." *The Review of Metaphysics* 43(1) (1989): 72–106.

———. *Self and World*. Oxford: Oxford University Press, 1999.

Casula, Mario. "A.G. Baumgarten entre G.W. Leibniz et Chr. Wolff." *Archives de Philosophie* 42 (1979): 547–574.

Clarke, Randolph. "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism." *Mind* 118 (470) (2009): 323–351.

Collins, Arthur. *Possible Experience*. Berkeley and Los Angeles: University of California Press, 1999.

Corr, Charles A. “Christian Wolff and Leibniz.” *Journal of the History of Ideas* 36(2) (1975): 241–262.

———. “Cartesian Themes in Wolff’s German Metaphysics.” In *Christian Wolff 1679–1754: Interpretationen zu seiner Philosophie und deren Wirkung*, edited by W. Schneiders, 113–20. Hamburg: Meiner, 1983.

Crusius, Christian August. *Die Philosophischen Hauptwerke*. 4 Volumes. Edited by G. Tonelli. Hildesheim, Olms, 1964–1987.

———. *Dissertatio philosophica de usu et limitibus principii rationis determinantis vulgo sufficientis*. Leipzig: 1743.

———. *Ausführliche Abhandlung von dem rechten Gebrauche und der Einschränkung des sogenannten Satzes vom zureichenen oder besser Determinierenden Grund*. Leipzig: 1744.

———. *Entwurf der nothwendigen Vernunft-Wahrheiten*. 1745. Leipzig: 1766.

Curley, Edwin. “Leibniz on Locke on Personal Identity.” In *Leibniz: Critical and Interpretive Essays*, edited by Michael Hooker, 302–326. Minneapolis: University of Minnesota Press, 1982.

Chalmers, David. *The Conscious Mind*. Oxford: Oxford University Press, 1996.

Davidson, Donald. “Mental Events” (1970). In *Essays on Actions and Events*. 2nd Edition. Oxford: Clarendon Press, 2001.

Descartes, René. *Oeuvres de Descartes*. 12 Volumes. Edited by C. Adam and P. Tannery. Paris: J. Vrin, 1964–1976.

———. *Discourse on Method and Meditations on First Philosophy*. Translated by Donald A. Cress. Indianapolis: Hackett, 1980.

———. *The Philosphical Writings of Descartes* and *Correspondence*. 3 Volumes. Edited and translated by J. Cottingham, R. Stoothoff, D. Murdoch, A. Kenny. Cambridge: Cambridge University Press, 1984–.

Dicker, Georges. *Kant’s Theory of Knowledge. An Analytical Introduction*. Oxford: Oxford University Press, 2004.

Downing, Lisa. “Locke’s Ontology.” In *The Cambridge Companion to Locke’s Essay*, edited by Lex Newman, 352–380. Cambridge: Cambridge University Press, 2007.

Dyck, Corey W. “Empirical Consciousness Explained: Self-Affection, (Self-)Consciousness and Perception in the B Deduction.” *Kantian Review* 11(2006): 29–54.

———. “The Subjective Deduction and the Search for a Fundamental Force.” *Kant-Studien* 99(2) (2008): 152–179.

———. “The Divorce of Reason and Experience: Kant's Paralogisms of Pure Reason in Context.” *Journal of the History of Philosophy* 47(2) (2009): 249–275.

———. “The Aeneas Argument: Personality and Immortality in Kant’s Third Paralogism.” *Kant Yearbook* 2 (2010): 95–122.

———. “A Wolff in Kant’s Clothing: Christian Wolff’s Influence on Kant’s Accounts of Consciousness, Self-Consciousness, and Psychology.” *Philosophy Compass* 6/1 (2011): 44–53.

———. *Kant and Rational Psychology*. Oxford: Oxford University Press, 2014.

Eberhard, Johann August. *Allgemeine Theorie des Denkens und Empfindens*. Berlin: 1776.

Emundts, Dina. “Die Paralogismen und die Widerlegung des Idealismus in Kants *Kritik der reinen Vernunft*.” *Deutsche Zeitschrift für Philosophie* 54(2) (2006): 295–309.

Ertl, Wolfgang. *Kants Auflösung der “dritten Antinomie”: Zur Bedeutung des Schöpfungskonzepts für die Freiheitslehre*. Freiburg: Karl Alber Verlag, 1998.

———. “Hud Hudson: *Kant's Compatibilism*.” *Kant-Studien* 90 (1999): 371–384.

———. “Schöpfung und Freiheit: Ein kosmologischer Schlüssel zu Kants Kompatibilismus.” In *Kants Metaphysik und Religionsphilosophie*, 43–76. Hamburg: Felix Meiner, 2004.

———. “‘Ludewig’ Molina and Kant’s Libertarian Compatibilism.” In *A Companion to Luis de Molina*, edited by A. Aichele and M. Kaufmann. Cologne: Brill, 2013.

Fara, Michael. “Masked Abilities and Compatibilism.” *Mind* 117(468) (2008): 843–865.

Forster, Michael N. *Kant and Skepticism*. Princeton: Princeton University Press, 2008.

Frankfurt, Harry. “Alternate Possibilities and Moral Responsibility.” *Journal of Philosophy* 66(23) (1969): 829–839.

Franks, Paul. *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism*. Cambridge: Harvard University Press, 2005.

Garber, Daniel. *Descartes’ Metaphysical Physics*. Chicago: University of Chicago Press, 1992.

Garrett, Brian. *Personal Identity and Self-Consciousness*. London: Routledge, 1998.

Grier, Michelle. “Illusion and Fallacy in Kant’s First Paralogism.” *Kant-Studien*, 84(3) (1993): 257–282.

———. *Kant’s Doctrine of Transcendental Illusion*, New York: Cambridge University Press, 2001.

Grüne, Stefanie. “Kant and the Spontaneity of the Understanding.” In *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, edited by Dina Emundts, 145–176. Berlin: Walter de Gruyter, 2013.

Guyer, Paul. “Kant’s Intentions in the Refutation of Idealism.” *Philosophical Review* 92(3) (1983): 329–383.

———. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press, 1987.

———. editor. *The Cambridge Companion to Kant*. Cambridge: Cambridge University Press, 1992.

———. *Kant*. New York: Routledge, 2006.

———. editor. *The Cambridge Companion to Kant and Modern Philosophy*. Cambridge: Cambridge University Press, 2007.

———. editor. *The Cambridge Companion to Kant’s Critique of Pure Reason*. Cambridge: Cambridge University Press, 2010.

Haakonssen, Knud, editor. *The Cambridge History of Eighteenth-Century Philosophy*. Cambridge: Cambridge University Press, 2006.

Hahmann, Andree. *Kritische Metaphysik der Substanz: Kant im Widerspruch zu Leibniz*. Berlin: Walter de Gruyter, 2009.

———. “Tetens über die Freiheit als Vermögen der Seele.” Forthcoming.

Hanna, Robert and A. Moore. “Reason, Freedom and Kant: An Exchange.” *Kantian Review* 12(1) (2007): 113–133.

Heimsoeth, Heinz. *Studien zur Philosophie Immanuel Kants*: *Metaphysische Ursprünge und Ontologische Grundlagen*. Köln: Kölner Universitäts Verlag, 1956.

———. *Transzendentale Dialektik, Dritter Teil*. Berlin: Walter de Gruyter, 1969.

Henrich, Dieter. “The Identity of the Subject in the Transcendental Deduction.” In *Reading Kant: New Perspectives on Transcendental Arguments and Critical Philosophy*, edited by Eva Schaper and Wilhelm Vossenkuhl, 250–280. Oxford: Blackwell, 1989.

———. *The Unity of Reason*. Edited by R.L. Velkley. Cambridge: Harvard University Press, 1994.

Heßbrüggen-Walter, Stefan. *Die Seele und ihre Vermögen: Kants Metaphysik des Mentalen in der Kritik der reinen Vernuft*. Paderborn: Mentis Verlag, 2004.

Hoeffe, Otfried. *Kants Kritik der reinen Vernunft: Die Grundlegung der modernen Philosophie*. München : C. H. Beck, 2003.

———. *Kant's Critique of Pure Reason: The Foundations of Modern Philosophy*. Dordrecht: Springer, 2010.

Hogan, Desmond. "How to Know Unknowable Things in Themselves." *Nous* 43(1) (2009): 49–63.

———. "Three Kinds of Rationalism and the non-Spatiality of Things in Themselves." *Journal of the History of Philosophy* 47(3) (2009): 355–82.

———. "Noumenal Affection." *Philosophical Review* 118(4) (2009): 501–532.

Hooker, Michael, editor. *Leibniz: Critical and Interpretive Essays*. Minneapolis: University of Minnesota Press, 1982.

Hoppe, Hansgeorg. *Synthesis bei Kant*. Berlin: de Gruyter, 1983

Horstmann, Rolf-Peter. "Kants Paralogismen." *Kant-Studien* 84(4) (1993): 408–425.

———. *Bausteine kritischer Philosophy: Arbeiten zu Kant*. Bodenheim: Philo, 1997.

Hudson, Hud. *Kant's Compatibilism*. Ithaca: Cornell University Press, 1994.

———. "Kant's Third Antinomy and Anomalous Monism." In *Immanuel Kant: Groundwork of the Metaphysics of Morals in Focus*, edited by Lawrence Pasternack, 234–267. London: Routledge, 2002.

Hume, David. *An Enquiry Concerning Human Understanding*. Edited by Tom L. Beauchamp. Oxford: Oxford University Press, 2000.

———. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. 3rd Edition. Edited by L.A. Selby-Bigge and P.H. Nidditch. Oxford: Oxford University Press, 1975.

———. *A Treatise of Human Nature*. Oxford: Oxford University Press, 2000.

Jacobi, Friedrich Heinrich. *David Hume über den Glauben; oder Idealismus und Realismus*. Breslau: 1787.

Kant, Immanuel *Kants gesammelte Schriften*. Edited by the Preussische Akademie der Wissenschaften and Deutsche Akademie der Wissenschaften zu Berlin. Berlin: De Gruyter, 1900–.

———. *Theoretical Philosophy 1755–1770*. Translated and Edited by D. Walford with R. Meerbote. Cambridge: Cambridge University Press, 1992.

———. *Lectures on Logic*. Edited and Translated by J. Michael Young. Cambridge: Cambridge University Press, 1992.

———. *Opus postumum*. Edited by E. Förster and translated by E. Förster and M. Rosen. Cambridge: Cambridge University Press, 1993.

———. *The Metaphysics of Morals*. Edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.

———. *Lectures on Metaphysics*. Edited and Translated by Karl Ameriks and Steve Naragon. Cambridge: Cambridge University Press, 1997.

———. *Critique of Pure Reason*. Edited by P. Guyer and A. Wood. Cambridge: Cambridge University Press, 1998.

———. *Religion within the Boundaries of Mere Reason*. Translated and edited by Allen Wood and George di Giovanni. Cambridge: Cambridge University Press, 1998.

———. *Correspondence*. Edited and Translated by A. Zweig. Cambridge: Cambridge University Press, 1999.

———. *Critique of the Power of Judgment*. Edited by Paul Guyer and translated by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press, 2000.

———. *Prolegomena to Any Future Metaphysics*. Translated by Gary Hatfield. Cambridge: Cambridge University Press, 2001.

———. *Kant's Theoretical Philosophy after 1781*. Edited by H. Allison, P. Heath and translated by G. Hatfield, M. Friedman, H. Allison, P. Heath. Cambridge: Cambridge University Press, 2002.

———. *Metaphysical Foundations of Natural Science*. Edited and translated by M. Friedman. Cambridge: Cambridge University Press, 2004.

———. *Notes and Fragments*. Edited by Paul Guyer and translated by Curtis Bowman, Paul Guyer, Frederick Rauscher. Cambridge: Cambridge University Press, 2005.

———. *Anthropology From A Pragmatic Point of View*. Edited and translated by Robert B. Louden. Cambridge: Cambridge University Press, 2006.

Keller, Pierre. *Kant and the Demands of Self-Consciousness*. Cambridge: Cambridge University Press, 1998.

Kemp Smith, Norman. *A Commentary to Kant's Critique of Pure Reason*. New York: Palgrave Macmillian, 2003.

Kim, Jaegwon. *Physicalism or Something Near Enough*. Princeton: Princeton University Press, 2005.

———. *Philosophy of Mind*. 2nd Edition. Cambridge: Cambridge University Press, 2006.

Kitcher, Patricia. "Kant on Self-Identity." *Philosophical Review* 91(1) (1982): 41–72.

———. "Kant's Paralogisms." *Philosophical Review* 91(4) (1982): 515–547.

———. *Kant's Transcendental Psychology*. Oxford: Oxford University Press, 1990.

———. *Kant's Thinker*. Oxford: Oxford University Press, 2011.

Klemme, Heiner F. *Kants Philosophie des Subjekts*. Hamburg: Felix Meiner Verlag, 1996.

Knutzen, Martin. *Philosophische Abhandlung von der immateriellen Natur der Seele.* Königsberg: 1744.

———. *Systema Causaurm Efficientium*. Leipzig: 1745.

Krüger, J.G. *Versuch einer Experimental-Seelenlehre*. Halle: 1756.

Korsgaard, Christine M. "Personal Identity and Unity of Agency." In *Creating the Kingdom of Ends*, 363–397. Cambridge: Cambridge University Press, 1996.

Lange, Joachim. *Modesta disquisitio novi philosophiae systematis de deo, mundo et homine*. Halle: 1723.

———. *Anmerckung über des Herrn […] Christian Wolffens Metaphysicam*. Kassel: 1724.

Langton, Rae. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Oxford University Press, 1998.

Laywine, Alison. Kant's *Early Metaphysics and the Origins of the Critical Philosophy*. Atascadero: Ridgeview Publishing Company, 1993.

———."Kant on the Self as Model of Experience." *Kantian Review* 9 (2005): 1–29.

———. "Kant's Metaphysical Reflections in the *Duisburg Nachlaß*." *Kant-Studien* 97 (2006): 79–113.

Lennon, Thomas M. and Stainton Robert J., editors. *The Achilles of Rationalist Psychology*. Dordrecht: Springer, 2008.

Leibniz, Gottfried Wilhelm. *New Essays on Human Understanding*. Edited by Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press, 1996.

———. *Philosophical Papers and Letters*. Second Edition. Edited and translated by Leroy E. Loemaker. Dordrecht: Kluwer Academic Publishers, 1989.

Longuenesse, Beatrice. *Kant and the Capacity to Judge*. Princeton: Princeton University Press, 1998.

———. "Kant's Deconstruction of the Principle of Sufficient Reason." *The Harvard Review of Philosophy* IX (2001): 67–87.

———. "Self-Consciousness and Consciousness of One's Own Body: Variations on a Kantian Theme." *Philosophical Topics* 34(1/2) (2006): 283–309.

———. “Kant on the Identity of Persons.” *Proceedings of the Aristotelian Society* 107(1) (2007): 149–67.

Locke, John. *An Essay Concerning Human Understanding*. Edited by Peter H. Nidditch, Oxford: Clarendon Press, 1975.

Ludwig, Bernd. “Die ‘*consequente Denkungsart* der speculativen Kritik.’ Kants radikale Umgestaltung seiner Freiheitslehre im Jahre 1786 und die Folgen für die Kritische Philosophie als Ganze.” *Deutsche Zeitschrift für Philosophie* 58(4) (2010): 595–628.

Mackie, J.L. *Problems from Locke*. Oxford: Oxford University Press, 1976.

Marshall, Colin, “Kant’s Metaphysics of the Self.” *Philosophers’ Imprint* 10(8) (2010): 1–21.

McCann, Edwin. “Locke on Identity: Matter, Life, and Consciousness.” *Archiv für Geschichte der Philosophie* 69(1) (1987): 54–77.

McLear, Colin. “Kant on Animal Consciousness.” *Philosophers’ Imprint* 11(15) (2011): 1–16.

McDowell, John. *Mind and World*. Cambridge: Harvard University Press, 1994.

Meerbote, Ralf. “The Unknowability of Things in Themselves.” In *Proceedings of the 3rd International Kant Congress*, edited by L. W. Beck, 415–423. Dordrecht: Springer, 1972.

———. “Kant on the Nondeterminate Character of Human Actions.” In *Kant on Causality, Freedom, and Objectivity*, edited by W. L Harper and Ralf Meerbote, 138–63. Minneapolis: University of Minnesota Press, 1984.

———. “Kant on Freedom and the Rational and Morally Good Will.” In *Self and Nature in Kant’s Philosophy*, edited by Allen Wood, 57–72. Ithaca: Cornell University Press, 1984.

Meier, Georg Friedrich. *Metaphysik*. Vol. 1. Halle: 1755.

Melnick, Arthur. *Kant’s Theory of the Self*. New York: Routledge, 2009.

Mendelssohn, Moses. *Phädon oder Über die Unsterblichkeit der Seele*. Berlin: 1767.

———. *Phaedon, or the Death of Socrates*. Translated by C. Cullen. London: 1789.

Noonan, Harold W. *Personal Identity*. Second edition. New York: Routledge, 2003.

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1986.

Pereboom, Derk. “Kant on Transcendental Freedom.” *Philosophy and Phenomenological Research* 73(3) (2006): 537–567.

Pippin, Robert. “Kant on the Spontaneity of the Mind.” In *Idealism as Modernism: Hegelian*

Variations, 29–55. Cambridge: Cambridge University Press, 1997.

Powell, C. Thomas. "Kant's Fourth Paralogism." *Philosophy and Phenomenological Research* 48(3) (1988): 389–414.

———. *Kant's Theory of Self-Consciousness*. Oxford: Oxford University Press, 1990.

Prauss, Gerold. *Erscheinungen bei Kant*. Berlin: de Gruyter, 1971.

———. *Kant und das Problem der Dinge an Sich*. Bonn: Bouvier, 1974.

Proops, Ian. "Kant's First Paralogism." *Philosophical Review* 119(4) (2010): 449–495.

Reid, Thomas. *Essays on the Intellectual Powers of Man*. London: 1785.

———. *Essays on The Powers of the Human Mind*. London: 1827.

Rosefeldt, Tobias. *Das logische Ich. Kant über den Gehalt des Begriffes von sich selbst*. Berlin: Philo Verlag, 2000.

———. "Kants Ich als Gegenstand." *Deutsch Zeitschrift für Philosophie* 54(2) (2006): 277–293.

———. "Kant's Self: Real Entity and Logical Identity." In *Strawson and Kant*, edited by Hans-Johann Glock, 141–154. Oxford: Oxford University Press, 2003.

———. "Dinge an sich und sekundäre Qualitäten." In *Kant in der Gegenwart*, edited by J. Stolzenburg, 167–209. Berlin: de Gruyter, 2007.

———. "Kants Kompatibilismus." In *Sind wir Bürger zweier Welten?: Freiheit und moralische Verantwortung im transzendentalen Idealismus*, edited by M. Brandhorst, A. Hahmann, B. Ludwig, 77–109. Hamburg: Felix Meiner Verlag, 2012.

Russell, Bertrand. *The Analysis of Matter*. London: Kegan Paul, 1927.

Sassen, Brigitte. "Kant and Mendelssohn on the Implications of the 'I think'." In *The Achilles of Rationalist Psychology*, edited by T.M. Lennon and R.J. Stainton, 215–233. Dordrecht: Springer, 2008.

Shoemaker, Sydney. "Persons and their Past." In *Identity, Cause and Mind*, 19–48. Cambridge: Cambridge University Press, 1984.

Schönfeld, Martin. *The Philosophy of the Young Kant: The Precritical Project*. Oxford: Oxford University Press, 2000.

Schulthess, Peter. *Relation und Funktion: Eine systematische entwicklungsgeschichtliche Untersuchung zur theoretischen Philosohie Kants*. Berlin: de Gruyter 1981.

Schulting, Dennis and Verburgt, Jacco, editors. *Kant's Idealism: New Interpretations of a Controversial Doctrine*. Dordrecht: Springer, 2011.

Schulting, Dennis. "Kant's Idealism: The Current Debate." In *Kant's Idealism: New Interpretations of a Controversial Doctrine*, edited by D. Schulting and J. Verburgt, 1–28. Dordrecht: Springer, 2011.

Sellars, Wilfrid. "…this I or He or It (The thing) which thinks…" *Proceedings and Addresses of the American Philosophical Association* 44 (1970/1971): 5–31.

Stang, Nicholas. "Who's Afraid of Double Affection." *Philosophers' Imprint* (forthcoming).

———. "The Non-Identity of Appearances and Things in Themselves." *Noûs* 47(4) (2013): 106–136.

———. "Kant on Complete Determination and Infinite Judgement." *British Journal for the History of Philosophy* 20(6) (2012): 1117–1139.

Strawson, P.F. *The Bounds of Sense*. London: Methuen, 1966.

Strawson, Galen. *Mental Reality*. Second edition. Cambridge: MIT Press, 2010.

Stubenberg, Leopold. *Consciousness and Qualia*. Philadelphia & Amsterdam: John Benjamins Publishers, 1998.

———. "Neutral Monism." *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition). Edited by Edward N. Zalta. URL= http://plato.stanford.eud/archives/spr2010/entrie/neutral-monism.*

Sturma, Dieter. *Kant über Selbstbewusstsein*. Hildesheim: Georg Olms Verlag, 1985.

Tester, Steven, editor and translator. *Georg Christoph Lichtenberg: Philosophical Writings*. Albany: State University of New York Press, 2012.

———. "G.C. Lichtenberg on Self-Consciousness and Personal Identity." *Archiv für Geschichte der Philosophie* 95(3) (2013): 336–359.

Tetens, Johann Nicolas. *Philosophische Versuche über die menschliche Natur und ihre Entwicklung*. Volume 2. Leipzig: 1777.

Thiel, Udo. *The Early Modern Subject: Self-consciousness and Personal Identity from Descartes to Hume*. Oxford: Oxford University Press, 2011.

Thümmig, Ludwig Philipp. *Institutiones philosophiae Wolfianae* I. Frankfurt: 1725.

Vaihinger, Hans, *Die Philosophie des Als Ob*. Leipzig: Felix Meiner Verlag, 1922.

———. *The Philosophy of 'As If': A System of the Theoretical, Practical and Religious Fictions of Mankind*. Translated by C. K. Ogden. London: Kegan Paul & Company, 1924.

Van Cleve, James. *Problems from Kant*. New York: Oxford University Press, 1999.

Van Inwagen, Peter. *An Essay on Free Will*. New York: Oxford University Press, 1983.

Vihvelin, Kadri. "Free Will Demystified: A Dispositional Account." *Philosophical Topics* 32(1/2) (2004): 427–450.

Vilhauer, Ben. "Can We Interpret Kant as a Compatibilist About Determinism and Moral Responsibility?" *British Journal for the History of Philosophy* 12(4) (2004): 719–730.

———. "Incompatibilism and Ontological Priority in Kant's Theory of Free Will." In *Rethinking Kant: Volume I*, edited by Pablo Muchnik, 22–47. Newcastle upon Tyne: Cambridge Scholars Publishing, 2008.

———. "The Scope of Responsibility in Kant's Theory of Free Will." *British Journal for the History of Philosophy* 18(1) (2010): 45–71.

Walker, Ralph C.S. *Kant*. London: Routledge & Kegan Paul, 1978.

Walsh, H.W. *Kant's Criticism of Metaphysics*. Edinburgh: Edinburgh University Press, 1997.

Watkins, Eric, editor and translator. *Kant's Critique of Pure Reason: Background Source Materials*. Cambridge: Cambridge University Press, 2009.

———. "Forces and Causes in Kant's Early Pre-Critical Writings." In *Studies in History and Philosophy of Science* 34 (2003): 5–27.

———. "Kant's Model of Causality: Causal Powers, Laws, and Kant's Reply to Hume." *Journal of the History of Philosophy* 42(4) (2004): 449–488.

———. *Kant and the Metaphysics of Causality*. Cambridge: Cambridge University Press, 2005.

Waxman, Wayne. *Kant's Model of the Mind*. New York: Oxford University Press, 1991.

Weinberg, Shelley. "The Metaphysical Fact of Consciousness in Locke's Theory of Personal Identity." *Journal of the History of Philosophy* 50(3) (2012): 387–345.

Whittle, Ann. "Dispositional Abilities." *Philosophers' Imprint* 10(12) (2010): 1–22.

Williams, Bernard. *Descartes: The Project of Pure Enquiry*. New York: Routledge, 1978.

Wilson, Margaret. "Leibniz: Self-Consciousness and Immortality in the Paris Notes and After." *Archiv für Geschichte der Philosophie* 58 (1976): 335–52.

Winkler, Kenneth. "Locke on Personal Identity." *Journal of the History of Philosophy* 29(2) (1991): 201–226.

Wolff, Christian. *Gesammelte Werke*. Edited by Jean École et al. Hildesheim: Georg Olms, 1962–.

———. *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* [*Deutsche Metaphysik*]. 1720. Hildesheim: George Olms, 1968.

———. *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt*. 1720. Halle: 1751.

———. *Psychologia empirica. Frankfurt: 1737. 1732. Reprinted Hildesheim: George Olms, 1968.*

———. *Psychologia rationalis*. Frankfurt: 1734. *Reprinted Hildesheim: George Olms, 1968.*

———. *Gesammelte Werke* III/23, J. Lange, *Kontroversschriften gegen die Wolffische Metaphysik*. Edited by Jean École. Hildesheim: Georg Olms, 1986.

———. *Gesammelte Werke* I/17, *Kleine Kotroversschriften mit Joachim Lange und Johann Franz Budde*. Edited by Jean École. Hildesheim: Georg Olms, 1980.

Wood, Allen W. *Kant's Rational Theology*. Ithaca: Cornell University Press, 1978.

Wood, Allen. "Kant's Compatibilism." In *Self and Nature in Kant's Philosophy*, edited by Allen Wood, 73–101. Ithaca: Cornell University Press, 1984.

Wunderlich, Falk. "Kant's Second Paralogism in Context." In *Between Leibniz, Newton and Kant*, edited by W. Lefevre, 175–188. Netherlands: Springer, 2001.

———. *Kant und die Bewußtseinstheorien des 18. Jahrhunderts*. Berlin: Walter de Gruyter, 2005.

Wuerth, Julian. "Kant's Immediatism–Pre-Critique." *Journal of the History of Philosophy* 44(4) (2006): 489–532.

———. "The Paralogisms of Pure Reason." In *The Cambridge Companion to Kant's Critique of Pure Reason*, edited by Paul Guyer, 210–244. Cambridge: Cambridge University Press, 2010.

———. "The First Paralogism, its Origin, and its Evolution: Kant on How the Soul Both Is and Is Not a Substance." In *Cultivating Personhood: Kant and Asian Philosophy*, edited by Stephen R. Palmquist, 157–166. Berlin: de Gruyter, 2010.

Wundt, Max. *Kant als Metaphysiker–Ein Beitrag zur Geschichte der deutschen Philosophie im 18. Jahrhundert*. Stuttgart: Ferdinand Enke, 1924.

Xie, Simon Shengjian. "What Is Kant: A Compatibilist or an Incompatibilist? A New Interpretation of Kant's Solution to the Free Will Problem." *Kant-Studien* 100 (2009): 53–76.

Yolton, John W. *Thinking Matter: Materialism in Eighteenth–Century Britain*. Oxford: Blackwell, 1984.

———. *Locke and French Materialism*. Oxford: Clarendon Press, 1991.

Ziedler, Johann Gottfried. *Pneumatica*. Halle: 1701.

Zuckert, Rachel. *Kant on Beauty and Biology: An Interpretation of the Critique of Judgment*. Cambridge: Cambridge University Press, 2007.